

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:14:04 ; Search time 47.36 seconds  
(without alignments) 292.164 Million cell updates/sec

Title: US-08-981-087a-2

Sequence: 1 SYTDKTLILYFNKLYKKIK.....LWYKTIWTLQDTAGNNOKL 144

Scoring table: OLIGO Gapop 60.0, Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: PIR.71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	100.0	366	2	neurotoxin type F
2	26	18.1	369	2	neurotoxin type F
3	26	18.1	1274	2	neurotoxin type F
4	15	10.4	1268	2	neurotoxin type F
5	11	7.6	367	2	neurotoxin type E
6	11	7.6	1251	2	neurotoxin type E
7	11	7.6	1252	2	neurotoxin type E
8	9	6.2	1296	2	neurotoxin type E
9	8	5.6	1072	2	neurotoxin type E
10	8	5.6	1291	1	neurotoxin type E
11	8	5.6	1291	1	neurotoxin type E
12	8	5.6	1291	1	neurotoxin type E
13	8	5.6	1291	1	neurotoxin type E
14	7	4.9	209	2	neurotoxin type E
15	7	4.9	209	2	neurotoxin type E
16	7	4.9	209	2	neurotoxin type E
17	7	4.9	209	2	neurotoxin type E
18	7	4.9	209	2	neurotoxin type E
19	7	4.9	209	2	neurotoxin type E
20	7	4.9	209	2	neurotoxin type E
21	7	4.9	209	2	neurotoxin type E
22	7	4.9	209	2	neurotoxin type E
23	7	4.9	209	2	neurotoxin type E
24	7	4.9	209	2	neurotoxin type E
25	7	4.9	209	2	neurotoxin type E
26	7	4.9	209	2	neurotoxin type E
27	7	4.9	209	2	neurotoxin type E
28	7	4.9	209	2	neurotoxin type E
29	7	4.9	209	2	neurotoxin type E

30	7	4.9	1010	2	T13167
31	7	4.9	1116	2	T16112
32	7	4.9	1711	2	C71625
33	7	4.9	1900	2	AG2391
34	7	4.9	2292	1	GNNYED
35	7	4.9	2292	1	GNNYED
36	7	4.9	2292	1	GNNYED
37	6	4.2	30	2	PD0006
38	6	4.2	56	2	S66323
39	6	4.2	57	2	S66314
40	6	4.2	73	2	T13199
41	6	4.2	97	2	T38991
42	6	4.2	97	2	A86846
43	6	4.2	99	2	G90113
44	6	4.2	101	2	S72281
45	6	4.2	112	2	C64498

## ALIGNMENTS

RESULT 1  
S48110  
neurotoxin type F - Clostridium botulinum (fragment)  
C:Species: Clostridium botulinum  
C:Date: 14-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S48110  
R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
A:Reference number: S48103; MUID:94013372  
A:Accession: S48110  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-366 <CAM>  
A:Cross-references: EMBL:X70821; NID:9407792; PIDN:CAA50152.1; PID:9407793  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 100.0%; Score 144; DB 2; Length 366;  
Best Local Similarity 100.0%; Pred. No. 4e-141;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYTDKTLILYFNKLYKKIKDMSITLDRYENKFTDISGYGNSISNGDVYIYSTNRNPF 60  
DB 214 SYTDKTLILYFNKLYKKIKDMSITLDRYENKFTDISGYGNSISNGDVYIYSTNRNPF 273  
QY 61 GYSSKPEVNIAQNDITLYNGRYONSISFVVRIPKFKYKVNINNEYIIDICIRNNNSG 120  
DB 274 GYSSKPEVNIAQNDITLYNGRYONSISFVVRIPKFKYKVNINNEYIIDICIRNNNSG 333  
QY 121 WKISLWYKTIWTLQDTAGNNOKL 144  
DB 334 WKISLWYKTIWTLQDTAGNNOKL 357  
RESULT 2  
S48109  
neurotoxin type F - Clostridium botulinum (fragment)  
C:Species: Clostridium botulinum  
C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 16-Jul-1999  
C:Accession: S48109  
R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
A:Reference number: S48103; MUID:94013372  
A:Accession: S48109  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-369 <CAM>  
A:Cross-references: EMBL:X70820; NID:9407790; PIDN:CAA50151.1; PID:9407791  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

C:Superfamily: tetanus toxin

Query Match 18.1%; Score 26; DB 2; Length 369;  
Best Local Similarity 100.0%; Pred. No. 5,6e-19;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 SILDREYNNKFDISGYSNISING 48  
DB 236 SILDREYNNKFDISGYSNISING 261

RESULT 3

I40813  
neurotoxin type F - Clostridium botulinum

C:Species: Clostridium botulinum

C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999

C:Accession: I40813; S48108

R:East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, P.T.

PEMS Microbiol. Lett. 96, 225-230, 1992

A:Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.

A:Reference number: I40644

A:Accession: I40813

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1274 <RES>

A:Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867

R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulin neurotoxin gene and specific id

A:Reference number: S48103; MUID:94013372

A:Accession: S48108

A>Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 634-1002 <CAM>

A:Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 18.1%; Score 26; DB 2; Length 1274;  
Best Local Similarity 100.0%; Pred. No. 1,6e-18;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 SILDREYNNKFDISGYSNISING 48  
DB 869 SILDREYNNKFDISGYSNISING 894

RESULT 4

S33411  
botulinum neurotoxin type F - Clostridium baratii

C:Species: Clostridium baratii

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999

C:Accession: S33411; S31860

R:Thompson, P.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.

PEMS Microbiol. Lett. 108, 173-182, 1993

A:Title: Nucleotide sequence of the gene coding for Clostridium baratii type F neurotoxin

A:Reference number: S33411; MUID:93252228

A:Accession: S33411

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1268 <THO>

A:Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 10.4%; Score 15; DB 2; Length 1268;  
Best Local Similarity 100.0%; Pred. No. 3,8e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 YISTNNQFGIYSS 65

DB 889 YISTNNQFGIYSS 903

RESULT 5

S48106  
neurotoxin type E - Clostridium botulinum (fragment)

C:Species: Clostridium botulinum

C:Date: 14-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999

C:Accession: S48106

R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulin neurotoxin gene and specific

A:Reference number: S48103; MUID:94013372

A:Accession: S48106

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-367 <CAM>

A:Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 7.6%; Score 11; DB 2; Length 367;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 NFSISFWVRIP 96  
DB 299 NFSISFWVRIP 309

RESULT 6

JH0256  
botulinum neurotoxin type E precursor - Clostridium butyricum

C:Species: Clostridium butyricum

C:Date: 30-Jun-1992 #sequence\_revision 15-May-1998 #text\_change 16-Jul-1999

C:Accession: JH0256; S16145

R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.

Biochem. Biophys. Res. Commun. 183, 107-113, 1992

A:Title: Sequences of the botulin neurotoxin E derived from Clostridium botulinum t

A:Reference number: JH0256; MUID:92181428

A:Accession: JH0256

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-27, 'E', '29-1251 <POU>

A:Cross-references: EMBL:X62088; NID:g40379

A:Experimental source: strains ATCC 43181 and ATCC 43755

R:Fujii, N.; Kimura, K.; Yashiki, T.; Inoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa,

J. Gen. Microbiol. 137, 519-525, 1991

A:Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E t

A:Reference number: S16145; MUID:91237316

A:Accession: S16145

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229, 'M', 231-252 <FUJ>

A:Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408

A:Experimental source: strain BL6340

C:Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter relea

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 7.6%; Score 11; DB 2; Length 1251;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 NFSISFWVRIP 96

Db 914 NFSISFWVRIP 924

## RESULT 7

botulinum neurotoxin type E precursor - Clostridium botulinum  
 C:Species: Clostridium botulinum  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Oct-1999  
 C:Accession: S21178; S48107; J0257; B35294; A60027; S18111  
 R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.  
 Eur. J. Biochem. 204, 657-667, 1992  
 A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin  
 A:Reference number: S21178; MUID:92174922  
 A:Accession: S21178  
 A:Molecule type: DNA  
 A:Residues: 1-1252 <MHE>  
 A:Cross-references: EMBL:X62683; NID:940397; PIDD:CAA4458.1; PID:940398  
 R:Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
 A:Reference number: S48103; MUID:94013372  
 A:Accession: S48107  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 616-982 <CAM>  
 A:Cross-references: EMBL:X70815; NID:9407786; PIDD:CAA50146.1; PID:9407787  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
 R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.  
 Biochem. Biophys. Res. Commun. 183, 107-113, 1992  
 A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type  
 A:Reference number: J0256; MUID:92181428  
 A:Accession: J0257  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-176, 'R', 178-197, 'C', 199-339, 'R', 341-772, 'I', 774-962, 'FE', 965-966, 'R', 968-1  
 A:Cross-references: EMBL:X62089; NID:940393; PIDD:CAA4399.1; PID:940394  
 A:Experimental source: strain Beluga  
 R:Binz, T.; Kuzanov, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.  
 J. Biol. Chem. 265, 9153-9158, 1990  
 A:Title: The complete sequence of botulinum neurotoxin type A and comparison with other  
 A:Reference number: A35294; MUID:90264400  
 A:Accession: B35294  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-176, 'R', 178-252 <BIN>  
 A:Experimental source: strain Beluga  
 R:Gienez, J.A.; Dasgupta, B.R.  
 Biochimie 72, 213-217, 1990  
 A:Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si  
 A:Reference number: A60027; MUID:90344918  
 A:Accession: A60027  
 A:Molecule type: protein  
 A:Residues: 420-427 <GIN>  
 A:Experimental source: strain Beluga  
 A:Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin  
 C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu  
 C:Superfamily: The heavy chain mediates the binding of toxin to cell receptors while the lig  
 C:Keywords: neurotoxin  
 F:422/2Product: botulinum neurotoxin type E light chain #status predicted <LCH>  
 F:423/2Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>  
 F:412-426/Disulfide bonds: #status predicted

Query Match 7.6%; Score 11; DB 2; Length 1252;  
 Best Local Similarity 100.0%; Pred. No. 0.0052;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 914 NFSISFWVRIP 924

## RESULT 8

botulinum neurotoxin type A - Clostridium botulinum  
 C:Species: Clostridium botulinum  
 C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 16-Jul-1999  
 C:Accession: I40645  
 R:Williams, A.; East, A.K.; Lawson, P.A.; Collins, M.D.  
 Res. Microbiol. 144, 547-556, 1993  
 A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type  
 A:Reference number: I40645; MUID:94143603  
 A:Accession: I40645  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1296 <RES>  
 A:Cross-references: EMBL:X73423; NID:9507070; PIDD:CAA51824.1; PID:9507071  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin

Query Match 6.2%; Score 9; DB 2; Length 1296;  
 Best Local Similarity 100.0%; Pred. No. 0.63;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 957 UNNETTII 965

## RESULT 9

hypothetical protein yqfG [imported] - Lactococcus lactis subsp. lactis (strain IL140  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: A86827  
 R:Botolin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Eh  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
 A:Reference number: A86825; MUID:21235186; PMID:11337471  
 A:Accession: A86827  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1072 <STO>  
 A:Cross-references: GB:A805176; PID:912724625; PIDD:AAK05715.1; GSPDB:GND00146  
 A:Experimental source: strain IL1403  
 C:Genetics: yqfG  
 A:Gene: yqfG

Query Match 5.6%; Score 8; DB 2; Length 1072;  
 Best Local Similarity 100.0%; Pred. No. 5.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 712 SSKRSEVN 719

## RESULT 10

botulinum neurotoxin type B (BoNT/B)  
 N:Alternate names: botulinum neurotoxin type B (BoNT/B)  
 C:Species: Clostridium botulinum  
 C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
 C:Accession: A48940; S21575; S21575; A42871; S07155; S08562; S07128; S08573; S08574  
 R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.  
 Appl. Environ. Microbiol. 58, 2345-2354, 1992  
 A:Title: Molecular cloning of the Clostridium botulinum structural gene encoding the  
 A:Reference number: A48940; MUID:92384550  
 A:Accession: A48940  
 A:Molecule type: DNA  
 A:Status: preliminary  
 A:Residues: 1-1291 <MHE>

A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735  
 A;Experimental source: type B, Danish  
 A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBI:P.112081); this publica  
 R;Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
 A;Reference number: S48103; MUID:94013372  
 A;Accession: S48105  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 634-994 <CAM>  
 A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783  
 A;Experimental source: proteolytic type B, strain NCTC 7273  
 R;Sabdo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.  
 submitted to the EMBL data library, April 1992  
 A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisid  
 A;Reference number: S21575  
 A;Accession: S21575  
 A;Molecule type: DNA  
 A;Residues: 36-217, 'G', 219-224, 'S', 226-246 <SZa>  
 A;Cross-references: EMBL:211934; NID:g40383; PIDN:CAA77991.1; PID:g40384  
 R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K  
 J. Biol. Chem. 267, 14721-14729, 1992  
 A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to  
 A;Reference number: A42871; MUID:92340505  
 A;Accession: A42871  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-313, 'S', 315-451 <KUR>  
 A;Experimental source: strain OKra  
 A;Note: sequence extracted from NCBI backbone (NCBI:P.109365)  
 R;Dasgupta, B.R.; Datta, A.  
 Biochimie 70, 811-817, 1988  
 A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with  
 A;Reference number: S07155; MUID:89000987  
 A;Accession: S07155  
 A;Molecule type: protein  
 A;Residues: 2-29, 'M', 31-45 <DAS>  
 A;Accession: S08562  
 A;Molecule type: protein  
 A;Residues: 442-463, 'R', 465-467 <DA2>  
 R;Schmidt, J.J.; Satyamorthy, V.; Dasgupta, B.R.  
 Arch. Biochem. Biophys. 238, 544-548, 1985  
 A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.  
 A;Reference number: S07128; MUID:85197963  
 A;Accession: S07128  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 2-16 <SCH1>  
 A;Accession: S08573  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 2-17 <SCH2>  
 A;Accession: S08574  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 442-459 <SCH3>  
 R;Schlavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; Dasgupta, B.R  
 Nature 359, 832-835, 1992  
 A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt  
 A;Reference number: S27125; MUID:93063293  
 A;Contents: annotation  
 C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap  
 C;Genetics:  
 A;Gene: bont/B  
 C;Function:  
 A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2  
 C;Superfamily: tetanus toxin  
 C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc  
 F;2-441/Product: botulinum B light chain #status experimental <LGHT>  
 F;442-1291/Product: botulinum B heavy chain #status experimental <HVT>  
 F;230, 234/Binding site: zinc (His) #status predicted  
 F;231/Active site: Glu #status predicted

Query Match 5.6%; Score 8; DB 1; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 8; Conservative 0; Mismatches .0; Indels 0; Gaps 0;  
 Oy 117 NNSGWRIS 124  
 |||||  
 Db 958 NNSGWRIS 965  
 RESULT 11  
 I40631  
 non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum  
 C;Species: Clostridium botulinum  
 C;Date: 12-Aug-1996 #sequence, revision 12-Aug-1996 #text\_change 16-Jul-1999  
 C;Accession: I40631; S48103; S48104; S36015  
 R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.  
 Curr. Microbiol. 28, 101-110, 1994  
 A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botul  
 A;Reference number: I40631; MUID:94122659  
 A;Accession: I40631  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1291 <RES>  
 A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149  
 R;Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A;Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
 A;Reference number: S48103; MUID:94013372  
 A;Accession: S48103  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 634-761, 'E', 763-841, 'M', 843, 'T', 845, 'N', 847-994 <CAM1>  
 A;Cross-references: EMBL:X70814; NID:g40778; PIDN:CAA50145.1; PID:g40779  
 A;Experimental source: non-proteolytic strain 2129B (Scott)  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
 A;Accession: S48104  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 634-843, 'T', 845, 'N', 847-994 <CAM2>  
 A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781  
 A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)  
 C;Comment: Botulinum neurotoxin type B in these strains may possess a capable catalyti  
 C;Genetics:  
 A;Gene: bont/B  
 C;Superfamily: tetanus toxin  
 C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc  
 F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>  
 F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVT>  
 F;230, 234/Binding site: zinc (His) #status predicted  
 F;231/Active site: Glu #status predicted  
 Query Match 5.6%; Score 8; DB 2; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 117 NNSGWRIS 124  
 |||||  
 Db 958 NNSGWRIS 965  
 RESULT 12  
 BTCLAB  
 botulinum B (EC 3.4.24.69) A precursor - Clostridium botulinum  
 N;Alternate names: botulinum neurotoxin type A  
 C;Species: Clostridium botulinum  
 C;Date: 31-Mar-1993 #sequence, revision 31-Mar-1993 #text\_change 18-Jun-1999  
 C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000  
 R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.  
 J. Biol. Chem. 265, 9153-9158, 1990  
 A;Title: The complete sequence of botulinum neurotoxin type A and comparison with oth



A:Reference number: A35294; MUID:90264400  
 A:Accession: A35294  
 A:Molecule type: DNA  
 A:Residues: 1-1296 <BIN>  
 A:Cross-references: GB:M30196; NID:q144864; PIDN:AAA23262.1; PID:q144865  
 A:Experimental source: strain 62A, subtype A  
 R:Thompson, D.E.; Brehm, J.K.; Oultrem, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.  
 Eur. J. Biochem. 189, 73-81, 1990  
 A:Title: The complete amino acid sequence of the clostridium botulinum type A neurotoxin  
 A:Reference number: S09492; MUID:90235864  
 A:Accession: S09492  
 A:Molecule type: DNA  
 A:Residues: 1, 'Q', '3-26, 'V', '28-1296 <THO>  
 A:Cross-references: EMBL:X52066; NID:q40381; PIDN:CAA36289.1; PID:q40382  
 A:Experimental source: NCTC 2916  
 R:Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.  
 FEBS Lett. 376, 41-44, 1995  
 A:Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components  
 A:Reference number: S67988; MUID:96096783  
 A:Accession: S68220  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-12 <FUJ>  
 A:Cross-references: EMBL:D67030; DDBJ:D50421; NID:q2160224  
 R:Beiley, M.J.; Somers, E.; Dasgupta, B.R.  
 Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989  
 A:Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term  
 A:Reference number: A33401; MUID:89350959  
 A:Accession: A33401  
 A:Molecule type: DNA  
 A:Residues: 1-35 <BET>  
 A:Cross-references: GB:M27892; NID:q144880; PIDN:AAA23269.1; PID:q551776  
 R:Simonez, J.A.; Dasgupta, B.R.  
 J. Protein Chem. 12, 351-363, 1993  
 A:Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and  
 A:Reference number: A53884; MUID:94000342  
 A:Accession: A53884  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 867-880:1148-1217, 'Y', '1219 <GIN>  
 A:Experimental source: strain Hall  
 A:Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex  
 R:Dasgupta, B.R.; Dekleva, M.L.  
 Biochimie 72, 661-664, 1990  
 A:Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and aro  
 A:Reference number: A60025; MUID:91120847  
 A:Accession: A60025  
 A:Molecule type: protein  
 A:Residues: 2-6,445-453, 'X', 455-457 <DAS1>  
 R:Dasgupta, B.R.; Foley, J.; Niece, R.  
 Biochemistry 26, 4162, 1987  
 A:Title: Partial sequence of the light chain of botulinum neurotoxin type A.  
 A:Reference number: A27000  
 A:Accession: A27000  
 A:Molecule type: protein  
 A:Residues: 2-47 <DAS2>  
 R:Blind, T.; Blaszi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;  
 J. Biol. Chem. 269, 1617-1620, 1994  
 A:Title: Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.  
 A:Reference number: A49708; MUID:94124495  
 A:Accession: A49708  
 A:Contents: annotation  
 C:Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap  
 C:Genetics:  
 A:Gene: atx, botA  
 C:Function:  
 C:Superfamily: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associa  
 C:Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot  
 F:444/Product: botoxilysin A light chain #status experimental <LGHT>  
 F:445-1296/Product: botoxilysin A heavy chain #status experimental <HVT>  
 F:223,227/Binding site: zinc (His) #status predicted  
 F:224/Active site: glu #status predicted

Query Match 5.6%; Score 8; DB 1; Length 1296;  
 Best Local Similarity 100.0%; Pred. No. 6.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 104 LNNEYTII 111  
 Db 958 LNNEYTII 965  
 RESULT 13  
 B70190  
 Conserved hypothetical protein B80723 - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence, revision 13-Feb-1998 #text, change 26-Aug-1999  
 C:Accession: B70190  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu  
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: B70190  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-177 <KLE>  
 A:Cross-references: GB:AE000783; TIGR:BB0723  
 A:Experimental source: strain B31  
 C:Superfamily: conserved hypothetical protein M0240

Query Match 4.9%; Score 7; DB 2; Length 177;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 14 KLYKKIK 20  
 Db 98 KLYKKIK 104

RESULT 14  
 G90445  
 Hypothetical protein S802711 [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence, revision 24-May-2001 #text, change 24-May-2001  
 C:Accession: G90445  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Ch  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder  
 arett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.  
 Submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: G90445  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-209 <KUR>  
 A:Cross-references: GB:AE006641; NID:q13816031; PIDN:AAK42822.1; GSPDB:GND0155  
 C:Genetics:  
 A:Gene: S802711

Query Match 4.9%; Score 7; DB 2; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 NKFTDIS 38  
 Db 53 NKFTDIS 59  
 RESULT 15  
 AC2379

hypothetical protein alr4587 [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AC2379  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2379  
A:status: preliminary  
A:molecule type: DNA  
A:Residues: 1-213 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAW6286.1; PID:g17133723; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr4587  
C:Superfamily: biopolymer transport protein exdB-1

Query Match 4.9%; Score 7; DB 2; Length 213;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 89 ISEWVRI 95  
|||  
Db 32 ISEWVRI 38

Search completed: August 15, 2002, 11:14:06  
Job time: 258 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:09:48 ; Search time 47.36 Seconds

(without alignments)  
874.462 Million cell updates/sec

Title: US-08-981-087a-1

Sequence: 431  
1 SYTDKILILYFNKLYKKIK.....TSSNGCFWSFKSHGQEN 431

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR-71.\*

1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	153	35.5	366	2 S48110	neurotoxin type F
2	27	6.3	1274	2 I40813	neurotoxin type F
3	26	6.0	369	2 S48109	neurotoxin type F
4	22	5.1	1268	2 S33411	botulinum neurotoxin
5	15	3.5	1251	2 JH0256	botulinum neurotoxin
6	15	3.5	1252	2 S21178	botulinum neurotoxin
7	11	2.6	367	2 S48106	neurotoxin type E
8	11	2.6	1296	1 BICL4B	botulinum neurotoxin
9	11	2.6	1296	2 I40645	botulinum neurotoxin
10	10	2.3	458	2 T02571	probable myosins
11	8	1.9	340	2 B87350	hypothetical prote
12	8	1.9	1072	2 A86827	hypothetical prote
13	8	1.9	1291	1 A48940	botulinum neurotoxin
14	8	1.9	1291	1 A40631	non-proteolytic bo
15	8	1.9	1297	2 S39791	neurotoxin - Clost
16	7	1.6	67	2 A11594	hypothetical prote
17	7	1.6	156	2 S20990	myosin regulatory
18	7	1.6	177	2 B70190	conserved hypotet
19	7	1.6	200	2 A81295	probable membrane
20	7	1.6	202	2 S20992	myosin regulatory
21	7	1.6	204	2 S22745	hypothetical prote
22	7	1.6	213	2 AC2379	hypothetical prote
23	7	1.6	213	2 AC2379	hypothetical prote
24	7	1.6	213	2 AC2379	hypothetical prote
25	7	1.6	213	2 AC2379	hypothetical prote
26	7	1.6	213	2 AC2379	hypothetical prote
27	7	1.6	213	2 AC2379	hypothetical prote
28	7	1.6	213	2 AC2379	hypothetical prote
29	7	1.6	213	2 AC2379	hypothetical prote

30	7	1.6	244	2 T28307	ORF MSV146 hypothe
31	7	1.6	259	2 APL544	RNA polymerase sig
32	7	1.6	259	2 APL186	RNA polymerase sig
33	7	1.6	261	2 G84453	probable GDSU-moti
34	7	1.6	264	2 C70105	pyridoxal kinase (
35	7	1.6	265	2 AC0814	cod(1)atamin adeno
36	7	1.6	267	2 G97450	ABC transporter, A
37	7	1.6	267	2 F81029	type II restrictio
38	7	1.6	267	2 B65021	mechado-Joseph dis
39	7	1.6	280	2 T47572	probable ATP-bind
40	7	1.6	307	2 C83188	hypothetical prote
41	7	1.6	330	2 E89791	hypothetical prote
42	7	1.6	347	2 T19989	hypothetical prote
43	7	1.6	358	2 S07594	hypothetical prote
44	7	1.6	375	2 A83636	hypothetical prote
45	7	1.6	379	2 B69344	hypothetical prote

#### ALIGNMENTS

RESULT 1  
S48110  
neurotoxin type F - Clostridium botulinum (fragment)  
C:Species: Clostridium botulinum  
C>Date: 14-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1995  
C:Accession: S48110  
R:Campbell, K.D.; Collins, M.D.; East, A.K.  
U: Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
A:Reference number: S48103; MID:94013372  
A:Accession: S48110  
A>Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-366 <CAM>  
A:Cross-references: EMBL:X70821; NID:9407792; PIDN:CAA50152.1; PID:9407793  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 35.5% Score 153; DB 2; Length 366;  
Best Local Similarity 100.0%; Pred. No. 6.6e-149;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTDKILILYFNKLYKKIKDINSIDPMRENNKFDISGYSNISTNGDYIYSNNNOF 60  
DB 214 SYTDKILILYFNKLYKKIKDINSIDPMRENNKFDISGYSNISTNGDYIYSNNNOF 273  
OY 61 GIYSKSPSVNIAQNDIYNGRYGFSISFWVRIPKRYKVLNNEYTIIDICIRNNNSG 120  
DB 274 GIYSKSPSVNIAQNDIYNGRYGFSISFWVRIPKRYKVLNNEYTIIDICIRNNNSG 333  
OY 121 WKISLNYKTIWTIQTDRAGNNOKLVNNTOMIS 153  
DB 334 WKISLNYKTIWTIQTDRAGNNOKLVNNTOMIS 366

RESULT 2  
I40813  
neurotoxin type F - Clostridium botulinum  
C:Species: Clostridium botulinum  
C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999  
C:Accession: I40813; S48108  
R:East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, FEMS Microbiol. Lett. 96, 225-230, 1992  
A:Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.  
A:Reference number: I40644  
A:Accession: I40813  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1274 <RBS>  
A:Cross-references: GB:M92906; NID:9144866; PIDN:AAA23263.1; PID:9144867  
R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993  
 A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
 A:Reference number: S48103; MUID:94013372  
 A:Accession: S48108  
 A:Status: preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 634-1002 <CAM>  
 A:Cross-references: EMBL:X70816; NID:9407788; PIDN:CAA50147.1; PID:9407789  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin

Query Match 6.3%; Score 27; DB 2; Length 1274;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-19;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 YINKMIFVITNNRLGNSRIRYINGNLI 183  
 Db 1006 YINKMIFVITNNRLGNSRIRYINGNLI 1032

RESULT 3  
 S48109  
 neurotoxin type F - Clostridium botulinum (fragment)  
 C:Species: Clostridium botulinum  
 C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 16-Jul-1999  
 C:Accession: S48109  
 R:Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
 A:Reference number: S48103; MUID:94013372  
 A:Accession: S48109  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-369 <CAM>  
 A:Cross-references: EMBL:X70820; NID:9407790; PIDN:CAA50151.1; PID:9407791  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
 C:Superfamily: tetanus toxin

Query Match 6.0%; Score 26; DB 2; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-18;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 SILDREYNNKFDISGYSNISTNG 48  
 Db 236 SILDREYNNKFDISGYSNISTNG 261

RESULT 4  
 S33411  
 botulinum neurotoxin type F - Clostridium barati  
 C:Species: Clostridium barati  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999  
 C:Accession: S33411; S31860  
 R:Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.  
 FEMS Microbiol. Lett. 108, 175-182, 1993  
 A:Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin  
 A:Reference number: S33411; MUID:93252228  
 A:Accession: S33411  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1268 <THO>  
 A:Cross-references: EMBL:X68262; NID:949138; PIDN:CAA48329.1; PID:949139  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin

Query Match 5.1%; Score 22; DB 2; Length 1268;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-14;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 198 DNLEKIVGNDTRVYGIIRYFK 219

Db 1036 DNLEKIVGNDTRVYGIIRYFK 1057

RESULT 5

JH0256  
 botulinum neurotoxin type E precursor - Clostridium butyricum  
 C:Species: Clostridium butyricum  
 C:Date: 30-Jun-1992 #sequence\_revision 15-May-1998 #text\_change 16-Jul-1999  
 C:Accession: JH0256; S16145  
 R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.  
 Biochem. Biophys. Res. Commun. 183, 107-113, 1992  
 A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum t  
 A:Reference number: JH0256; MUID:92181428  
 A:Accession: JH0256  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA

A:Residues: 1-27, 'E', 29-1251 <POU>  
 A:Cross-references: EMBL:X62088; NID:940379  
 A:Experimental source: strains ATCC 43181 and ATCC 43755  
 R:Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, J.  
 Gen. Microbiol. 137, 519-525, 1991  
 A:Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E t  
 A:Reference number: S16145; MUID:91237316  
 A:Accession: S16145  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-229, 'M', 231-252 <FUJ>  
 A:Cross-references: EMBL:X5180; NID:940407; PIDN:CAA37321.1; PID:940408  
 A:Experimental source: strain B16340  
 C:Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter relea  
 C:Comment: The heavy chain mediates the binding of toxin to cell receptors while the  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin  
 F:2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>  
 F:423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEX>  
 F:412-426/Dissulfide bonds: #status predicted

Query Match 3.5%; Score 15; DB 2; Length 1251;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 ISDYINKMIFVITTN 168  
 Db 983 ISDYINKMIFVITTN 997

RESULT 6  
 S21178  
 botulinum neurotoxin type E precursor - Clostridium botulinum  
 C:Species: Clostridium botulinum  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Oct-1999  
 C:Accession: S21178; S48107; JH0257; B35294; A60027; S18111  
 R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.D.; Atkinson, T.; Minton, N.P.  
 Eur. J. Biochem. 204, 657-667, 1992  
 A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neuroto  
 A:Reference number: S21178; MUID:92174922  
 A:Accession: S21178  
 A:Molecule type: DNA  
 A:Residues: 1-1252 <RHE>  
 A:Cross-references: EMBL:X62683; NID:940397; PIDN:CAA44558.1; PID:940398  
 R:Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
 A:Reference number: S48103; MUID:94013372  
 A:Accession: S48107  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 616-982 <CAM>  
 A:Cross-references: EMBL:X70815; NID:9407786; PIDN:CAA50146.1; PID:9407787  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
 R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.

Biochem. Biophys. Res. Commun. 183, 107-113, 1992  
 A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type  
 A:Reference number: JH0256; MUID:92181428  
 A:Accession: JH0257  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-176, 'R', 178-252 <BIN>  
 A:Experimental source: strain Beluga  
 R:Gimenez, J.A.; Dasgupta, B.R.  
 Biochimie 72, 213-217, 1990  
 A:Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the s  
 A:Reference number: A60027; MUID:90344918  
 A:Accession: A60027  
 A:Molecule type: protein  
 A:Residues: 420-427 <GIN>  
 A:Experimental source: strain Beluga  
 A:Note: This fragment was generated by proteolysis with Lys-C rather than with trypsin  
 C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin  
 F:1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>  
 F:423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>  
 F:412-426/Disulfide bonds: #status predicted

Query Match 3.5%; Score 15; DB 2; Length 1252;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 ISDYINKMIFWTIN 168  
 ||||||||||||||||  
 Db 983 ISDYINKMIFWTIN 997

RESULT 7  
 S48106  
 neurotoxin type E - Clostridium botulinum (fragment)  
 C:Species: Clostridium botulinum  
 C:Date: 14-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
 C:Accession: S48106  
 R:Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
 A:Reference number: S48103; MUID:94013372  
 A:Accession: S48106  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-367 <CAM>  
 A:Cross-references: EMBL:X70818; NID:9407784; PIDN:CAA50149.1; PID:9407785  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin

Query Match 2.6%; Score 11; DB 2; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 0.0058;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 NFSISFWAIP 96  
 ||||||||||||  
 Db 299 NFSISFWAIP 309

RESULT 8  
 BNCIAB  
 nontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum  
 C:Alternate names: botulinum neurotoxin type A  
 C:Species: Clostridium botulinum  
 C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 18-Jun-1999  
 C:Accession: A35294; S09492; S68220; A33401; A33884; A60025; A27000  
 R:Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.  
 J. Biol. Chem. 265, 9153-9158, 1990  
 A:Title: The complete sequence of botulinum neurotoxin type A and comparison with oth  
 A:Reference number: A35294; MUID:90264400  
 A:Accession: A35294  
 A:Molecule type: DNA  
 A:Residues: 1-1296 <BIN>  
 A:Cross-references: GB:M30196; NID:9144864; PIDN:AAA23262.1; PID:9144865  
 A:Experimental source: strain 62A, subtype A  
 R:Thompson, D.E.; Brehm, J.K.; Oulttram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson,  
 Eur. J. Biochem. 189, 73-81, 1990  
 A:Title: The complete amino acid sequence of the Clostridium botulinum type A neuroto  
 A:Reference number: S09492; MUID:90235864  
 A:Accession: S09492  
 A:Molecule type: DNA  
 A:Residues: 1, 'Q', 3-26, 'V', 28-1296 <THO>  
 A:Cross-references: EMBL:X52066; NID:940381; PIDN:CAA36289.1; PID:940382  
 A:Experimental source: NCTC 2916  
 R:Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.  
 FEBS Lett. 376, 41-44, 1995  
 A:Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin compone  
 A:Reference number: S67988; MUID:96096783  
 A:Accession: S68220  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-12 <FUJ>  
 A:Cross-references: EMBL:D67030; DDBJ:D50421; NID:92160224  
 R:Belley, M.J.; Somers, E.; Dasgupta, B.R.  
 Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989  
 A:Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-ter  
 A:Reference number: A33401; MUID:93305959  
 A:Accession: A33401  
 A:Molecule type: DNA  
 A:Residues: 1-35 <BRT>  
 A:Cross-references: GB:M27892; NID:9144880; PIDN:AAA23269.1; PID:9551776  
 R:Gimenez, J.A.; Dasgupta, B.R.  
 J. Proteins, Chem. 12, 351-363, 1993  
 A:Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42,  
 A:Reference number: A33884; MUID:94000342  
 A:Accession: A33884  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 867-880; 1148-1217, 'Y', 1219 <GIN>  
 A:Experimental source: strain Hall  
 A:Note: Sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after  
 R:Dasgupta, B.R.; Dekleva, M.L.  
 Biochimie 72, 661-664, 1990  
 A:Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and e  
 A:Reference number: A60025; MUID:91120847  
 A:Accession: A60025  
 A:Molecule type: protein  
 A:Residues: 2-6; 445-453, 'X', 455-457 <DAS1>  
 R:Dasgupta, B.R.; Foley, J.; Niece, R.  
 Biochemistry 26, 4162, 1987  
 A:Title: Partial sequence of the light chain of botulinum neurotoxin type A.  
 A:Reference number: A27000  
 A:Accession: A27000  
 A:Molecule type: protein  
 A:Residues: 2-47 <DAS2>  
 R:Binz, T.; Blaszi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.  
 J. Biol. Chem. 269, 1617-1620, 1994  
 A:Title: Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.  
 A:Reference number: A49708; MUID:94124495  
 A:Contents: annotation  
 C:Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic sy

C:Genetics:  
A:Gene: atx; bota  
C:Function:  
A:Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate  
C:Superfamily: tetanus toxin  
C:Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot  
F:2-444/Product: botulysin A light chain #status experimental <LGHT>  
F:445-1296/Product: botulysin A heavy chain #status experimental <HVV>  
F:223,227/Binding site: zinc (His) #status predicted  
F:224/Active site: Glu #status predicted

Query Match 2.6%; Score 11; DB 1; Length 1296;  
Best Local Similarity 100.0%; Pred. No. 0.019;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 WIEVTITNNRL 171  
|||||  
Db 1014 WIEVTITNNRL 1024

## RESULT 9

140645

botulinum neurotoxin type A - Clostridium botulinum

C:Species: Clostridium botulinum

C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 16-Jul-1999

C:Accession: 140645

R:Williams, A.; East, A.K.; Lawson, P.A.; Collins, M.D.

Res. Microbiol. 144, 547-556, 1993

A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A

A:Reference number: 140645; MUID:94143603

A:Accession: 140645

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1296 &lt;RES&gt;

A:Cross-references: EMBL:X73423; NID:g507070; PIDN:CA51824.1; PID:g507071

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 2.6%; Score 11; DB 2; Length 1296;  
Best Local Similarity 100.0%; Pred. No. 0.019;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 WIEVTITNNRL 171  
|||||  
Db 1014 WIEVTITNNRL 1024

## RESULT 10

102571

Probable myrosinase-binding protein [Imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein T16B24.5

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001

C:Accession: 102571; G84815

R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,

submitted to the EMBL Data Library, August 1998

A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.

A:Reference number: 214679

A:Accession: T02571

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-458 &lt;ROU&gt;

A:Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402676

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Mofrac, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umayam, L.; Tallon, L.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: AB4420; MUID:20083487

A:Accession: G84815

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-458 <STO>  
A:Cross-references: GB:AE002093; NID:g3402676; PIDN:AAC28979.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: T16B24.5; At2g39310  
A:Map position: 2  
A:Introns: 67/3; 221/3; 374/3

Query Match 2.3%; Score 10; DB 2; Length 458;  
Best Local Similarity 100.0%; Pred. No. 0.076;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 285 YOKPNIFSNT 294  
|||||  
Db 238 YOKPNIFSNT 247

## RESULT 11

B87350

hypothetical protein CC0813 [Imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: B87350

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Lauv, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87350

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-540 &lt;STO&gt;

A:Cross-references: GB:AE005673; NID:g13422062; PIDN:AAK22798.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0813

Query Match 1.9%; Score 8; DB 2; Length 540;  
Best Local Similarity 100.0%; Pred. No. 10;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 238 PDPSTIKD 245  
|||||  
Db 71 PDPSTIKD 78

## RESULT 12

A86827

Hypothetical protein yqfG [Imported] - Lactococcus lactis subsp. lactis (strain IL140

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: A86827

R:Boletun, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Eh

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: A86827

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1072 &lt;STO&gt;

A:Cross-references: GB:AE005176; PID:g12724625; PIDN:AAK05715.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: yqfG

Query Match 1.9%; Score 8; DB 2; Length 1072;  
Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 SSKPSEVN 71  
 |||||  
 Db 712 SSKPSEVN 719  
  
 RESULT 13  
 A48940  
 bontolixysin (EC 3.4.24.69) B precursor - Clostridium botulinum  
 N:Alternate names: botulinum neurotoxin type B (BoNT/B)  
 C:Species: Clostridium botulinum  
 C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
 C:Accession: A48940; S48103; S21575; S07155; S08362; S07128; S08573; S08574  
 R:McLellan, S.M.; Emswiler, M.J.; Bodsworth, N.J.; Breilm, J.K.; Atkinson, T.; Mincon, N.P.  
 Appl. Environ. Microbiol. 58, 2345-2354, 1992  
 A:Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type B  
 A:Reference number: A48940; MUID:92384550  
 A:Accession: A48940  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1291 <WHE>  
 A:Cross-references: GB:M81186; NID:944734; PIDN:AAA2321.1; PID:9444735  
 A:Experimental source: type B, Danish  
 A:Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica  
 R:Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
 A:Reference number: S48103; MUID:94013372  
 A:Accession: S48103  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 634-994 <CAM>  
 A:Cross-references: EMBL:X70817; NID:9407782; PIDN:CAA50148.1; PID:9407783  
 R:Sabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.  
 submitted to the EMBL Data Library, April 1992  
 A:Description: Partial amino acid sequence of botulinum neurotoxin type B and comparativ  
 A:Reference number: S21575  
 A:Accession: S21575  
 A:Molecule type: DNA  
 A:Residues: 36-217, 'G', 219-224, 'S', 226-246 <SZ>  
 A:Cross-references: EMBL:211934; NID:940383; PIDN:CAA7799.1; PID:940384  
 R:Kunzono, H.; Mochida, S.; Bittz, T.; Elsel, U.; Quanz, M.; Grebenstein, O.; Wenmans, K  
 J. Biol. Chem. 267, 14721-14729, 1992  
 A:Title: Minimal essential domains specifying toxicity of the light chains of tetanus to  
 A:Reference number: A42871; MUID:92340509  
 A:Accession: A42871  
 A:Status: preliminary  
 A:Molecule type: nucleic acid sequence not shown  
 A:Residues: 1-313, 'S', 315-451 <KUR>  
 A:Experimental source: strain OKra  
 A:Note: sequence extracted from NCBI backbone (NCBIP:109365)  
 R:Dasgupta, B.R.; Datta, A.  
 Biochimie 70, 811-817, 1988  
 A:Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with  
 A:Reference number: S07155; MUID:89000987  
 A:Accession: S07155  
 A:Molecule type: protein  
 A:Residues: 2-29, 'M', 31-45 <DAS>  
 A:Accession: S08562  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 442-463, 'R', 465-467 <DA2>  
 R:Schmidt, J.J.; Sathyanarayanan, V.; Dasgupta, B.R.  
 Arch. Biochem. Biophys. 238, 544-548, 1985  
 A:Title: Partial amino acid sequences of botulinum neurotoxins types B and E.  
 A:Reference number: S07128; MUID:85197963  
 A:Accession: S07128  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 2-16 <SCH1>  
 A:Accession: S08573  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 2-17 <SCH2>

A:Accession: S08574  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 442-459 <SCH3>  
 R:Schivo, G.; Benfante, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; Dasgupta,  
 Nature 359, 832-835, 1992  
 A:Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteo  
 A:Reference number: S27125; MUID:93063293  
 A:Accession: S27125  
 A:Content: annotation  
 C:Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic sy  
 C:Genetics:  
 A:Gene: bont/b  
 C:Function:  
 A:Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2  
 C:Superfamily: tetanus toxin  
 C:Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc  
 F:2-441/Product: bontolixysin B light chain #status experimental <LGHT>  
 F:442-1291/Product: bontolixysin B heavy chain #status experimental <HVT>  
 F:230,234/Binding site: zinc (His) #status predicted  
 F:231/Active site: Glu #status predicted  
  
 Query Match 1.9%; Score 8; DB 1; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 117 NNSGWRIS 124  
 |||||  
 Db 958 NNSGWRIS 965  
  
 RESULT 14  
 A40631  
 non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum  
 C:Species: Clostridium botulinum  
 C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 16-Jul-1999  
 C:Accession: I40631; S48104; S36015  
 R:Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.  
 Curr. Microbiol. 28, 101-110, 1994  
 A:Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botu  
 A:Reference number: I40631; MUID:94122659  
 A:Accession: I40631  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1291 <RES>  
 A:Cross-references: EMBL:X71343; NID:9296148; PIDN:CAA50482.1; PID:9296149  
 R:Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
 A:Reference number: S48103; MUID:94013372  
 A:Accession: S48103  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 634-761, 'E', 763-841, 'M', 843, 'T', 845, 'N', 847-994 <CAM1>  
 A:Cross-references: EMBL:X70814; NID:9407778; PIDN:CAA50145.1; PID:9407779  
 A:Experimental source: non-proteolytic strain 2129b (Scott)  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
 A:Accession: S48104  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 634-843, 'T', 845, 'N', 847-994 <CAM2>  
 A:Cross-references: EMBL:X70819; NID:9407780; PIDN:CAA50150.1; PID:9407781  
 A:Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)  
 C:Comment: Botulinum neurotoxin type B in these strains may possess a capable catalyti  
 C:Genetics:  
 A:Gene: bont/b  
 C:Superfamily: tetanus toxin  
 C:Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc  
 F:2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>  
 F:442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVT>  
 F:230,234/Binding site: zinc (His) #status predicted  
 F:231/Active site: Glu #status predicted

Query Match 1.9%; Score 8; DB 2; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 117 NNSGWRIS 124  
 |||||  
 DB 958 NNSGWRIS 965

RESULT 15  
 S39791  
 neurotoxin - Clostridium botulinum  
 C:Species: Clostridium botulinum  
 C:Date: 07-Oct-1994 #sequence\_revision 01-Dec-1995 #text\_change 16-Jul-1999  
 C:Accession: S39791  
 R:Campbell, K.; Collins, M.D.; East, A.K.  
 Biochim. Biophys. Acta 1216, 487-491, 1993  
 A:Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a  
 A:Reference number: S39791; MUID:94092745  
 A:Accession: S39791  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1297 <CAM>  
 A:Cross-references: EMBL:X74162; NID:g441275; PIDN:CAAS2275.1; PID:g441276  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin

Query Match 1.9%; Score 8; DB 2; Length 1297;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 154 ISDYINKW 161  
 |||||  
 DB 1002 ISDYINKW 1009

Search completed: August 15, 2002, 11:14:04  
 Job time: 256 sec





CC A polypeptide (AAW09014) comprises the heavy chain (amino acids  
 CC 848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be  
 CC produced using a synthetic gene (AAT48101) based on the natural  
 CC gene sequence (AAT48100) for the heavy chain. The polypeptides and  
 CC its fragments (see also AAW09015-17) lack the light chain and HN  
 CC epitopes necessary for metalloprotease activity and toxin  
 CC internalisation. They are free of botulinum toxin activity but can  
 CC induce protective immunity to a type F botulinum toxin, making them  
 CC useful for vaccine prodn. Recombinant polypeptides can be  
 CC produced in transformed host cells, esp. as fusion proteins, e.g.  
 CC with maltose binding protein to facilitate purification.

XX Sequence 431 AA;

Query Match 100.0%; Score 431; DB 18; Length 431;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNKILILYFNKLYKKIKDMSIDMRYNKFFIDISGYSNISINGDVIYSTNRNQF 60  
 Db 1 sytnklllyfnklykkikdmsldmryenkffidisgysnisngdvylystnrgf 60

QY 61 GIYSKPESEVNTAQNNDIYNGRYONFSEFWRIPEKYNKYNLNNEYTIIDCIRNNNSG 120  
 Db 61 g1yskspesevnlaqnddilyngryqnfswrlpkyfnkynlnneytlidcirmnsg 120

QY 121 WKISLNYKRIITWLODFGANNOKLVFNNTOMISIDYIKKIFVTITNRLGNSRIYNG 180  
 Db 121 wkislnykriitwldofgannoklvfnntomisidyikkfvtitnrlgnsriyng 180

QY 121 WKISLNYKRIITWLODFGANNOKLVFNNTOMISIDYIKKIFVTITNRLGNSRIYNG 180  
 Db 121 wkislnykriitwldofgannoklvfnntomisidyikkfvtitnrlgnsriyng 180

QY 181 NLIDEKTSISNLGDIHVSNDILFKIVGCDNTRYVGRIRKVPDELTGKTEIEFLYSDEPP 240  
 Db 181 nlidektsisnlgdihvsndilfkivgcdntryvgirrkvpdeltgkteieflysdepp 240

QY 241 SLIKDFGNYLLYKRYLLNLRTDKSITONSFLINQOQGVYOKPIESNTFLYTCV 300  
 Db 241 slikdfgnyllykryyllnlrtksitqnsflninqgvyqkpiisntfltycv 300

QY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINVDREYRLYADISIAKPEKIIKIRTSNSN 360  
 Db 301 eviirngstdisntdnfvrkndlayinvdreyrlyadisiakpekiklirtsnsn 360

QY 361 NSLGOIYWDISGNCTNPFONNNGCNTGLGFHSNNLVASWYNNIRKNTSNGCEWS 420  
 Db 361 nslgdiywdisgnctnfmfnngngnlgfghsnlvasswyynlirktltsngcfs 420

QY 421 FISKEGMOEN 431  
 Db 421 fiskegwgen 431

RESULT 2  
 AAB04096  
 ID AAB04096 standard; Protein; 432 AA.

XX AAB04096;  
 XX 11-Apr-2001 (first entry)

XX Botulinum toxin heavy chain C-terminal sequence (serotype F) ..

XX Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;  
 KW recombinant vector; antigen; immune response; vaccine; bacterium;  
 infection.

XX Synthetic.  
 OS Clostridium botulinum.

XX WO200067700-A2.  
 PD 16-NOV-2000.

PF 12-MAY-2000; 2000WO-US12890.  
 XX 12-MAY-1999; 99US-0133865.  
 PR 12-MAY-1999; 99US-0133866.  
 PR 12-MAY-1999; 99US-0133867.  
 PR 12-MAY-1999; 99US-0133868.  
 PR 12-MAY-1999; 99US-0133869.  
 PR 12-MAY-1999; 99US-0133873.  
 PR 29-JUL-1999; 99US-0146192.  
 XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
 XX WPI; 2001-016048/02.  
 DR N-PSDB; AAA54490.

PT New nucleic acids encoding the carboxy- or amino-terminal portions of  
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 PT vaccine against botulinism

XX Claim 3; Fig 9b; 73pp; English.

CC Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 CC chain and then posttranslationally nicked, forming a dchain  
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 CC remain linked by a disulfide bond. Nucleic acids encoding the  
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 CC expression vectors and expressed in transformed cells to produce  
 CC peptide antigens useful for eliciting an immune response to give  
 CC protective immunity against botulinum neurotoxin, which causes  
 CC botulism. The nucleic acids are expressible in a recombinant  
 CC organisms such as Escherichia coli or Pichia pastoris. The use  
 CC of recombinant nucleic acids are advantageous since it eliminates  
 CC the need to culture large quantities of hazardous toxin-producing  
 CC bacterium. Production yield from the genetically engineered product  
 CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.

XX Sequence 432 AA;

Query Match 100.0%; Score 431; DB 22; Length 432;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNKILILYFNKLYKKIKDMSIDMRYNKFFIDISGYSNISINGDVIYSTNRNQF 60  
 Db 2 sytnklllyfnklykkikdmsldmryenkffidisgysnisngdvylystnrgf 61

QY 61 GIYSKPESEVNTAQNNDIYNGRYONFSEFWRIPEKYNKYNLNNEYTIIDCIRNNNSG 120  
 Db 62 g1yskspesevnlaqnddilyngryqnfswrlpkyfnkynlnneytlidcirmnsg 121

QY 121 WKISLNYKRIITWLODFGANNOKLVFNNTOMISIDYIKKIFVTITNRLGNSRIYNG 180  
 Db 122 wkislnykriitwldofgannoklvfnntomisidyikkfvtitnrlgnsriyng 181

QY 181 NLIDEKTSISNLGDIHVSNDILFKIVGCDNTRYVGRIRKVPDELTGKTEIEFLYSDEPP 240  
 Db 182 nlidektsisnlgdihvsndilfkivgcdntryvgirrkvpdeltgkteieflysdepp 241

QY 241 SLIKDFGNYLLYKRYLLNLRTDKSITONSFLINQOQGVYOKPIESNTFLYTCV 300  
 Db 242 slikdfgnyllykryyllnlrtksitqnsflninqgvyqkpiisntfltycv 301

QY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINVDREYRLYADISIAKPEKIIKIRTSNSN 360  
 Db 302 eviirngstdisntdnfvrkndlayinvdreyrlyadisiakpekiklirtsnsn 361

QY 361 NSLGOIYWDISGNCTNPFONNNGCNTGLGFHSNNLVASWYNNIRKNTSNGCEWS 420  
 Db 362 nslgdiywdisgnctnfmfnngngnlgfghsnlvasswyynlirktltsngcfs 420

DB 362 ns1gqilvmdsigmctmfnngngnig1lghfshnlvasewymnrkntsgcfs 421  
 QY 421 FISKHEGMOEN 431  
 |||||||  
 DB 422 fiskhegwgen 432

RESULT 3  
 AAB04103  
 ID AAB04103 standard; Protein: 432 AA.  
 AC AAB04103;  
 XX  
 XX 11-APR-2001 (first entry)  
 DE Botulinum toxin heavy chain C-terminal sequence (serotype F).  
 XX Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;  
 KW recombinant vector; antigen; immune response; vaccine; bacterium;  
 KM infection.  
 XX Synthetic.  
 OS Clostridium botulinum.  
 XX  
 XX WO200067700-A2.  
 PN 16-NOV-2000.  
 XX  
 XX 12-MAY-2000; 2000MO-US12890.  
 PF  
 XX 12-MAY-1999; 99US-0133865.  
 PR 12-MAY-1999; 99US-0133866.  
 PR 12-MAY-1999; 99US-0133867.  
 PR 12-MAY-1999; 99US-0133868.  
 PR 12-MAY-1999; 99US-0133869.  
 PR 12-MAY-1999; 99US-0133873.  
 PR 29-JUL-1999; 99US-0146192.  
 XX  
 XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 PA  
 XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
 PI  
 XX WPI: 2001-016048/02.  
 DR N-PSDB; AAA54499.  
 DR  
 XX  
 PT New nucleic acids encoding the carboxy- or amino-terminal portions of  
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 PT vaccine against botulinism  
 PS  
 XX Disclosure: Fig 18b; 73pp; English.  
 PS  
 XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 CC chain and then posttranslationally nicked, forming a dichain  
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 CC remain linked by a disulfide bond. Nucleic acids encoding the  
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 CC expression vectors and expressed in transformed cells to produce  
 CC peptide antigens useful for eliciting an immune response to give  
 CC protective immunity against botulinum neurotoxin, which causes  
 CC botulism. The nucleic acids are expressible in a recombinant  
 CC organisms such as Escherichia coli or Pichia pastoris. The use  
 CC of recombinant nucleic acids are advantageous since it eliminates  
 CC the need to culture large quantities of hazardous toxin-producing  
 CC bacterium. Production yield from the genetically engineered product  
 CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.  
 XX  
 XX Sequence 432 AA;  
 SO

Query Match 100.0%; Score 431; DB 22; Length 432;  
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SYTNDKILILYFNKLYKIKDNDIRRYENKKEFIDISGYSNLSINGDVYIYSTRNOF 60  
 |||||||  
 DB 2 eytdnkallilyfnklykkikdndldmryenkffidlsygsnlsingdvlysturngf 61  
 |||||||

QY 61 GYSSKSPSEVNIQNDNDIYNGRYONPSISFWRIIPKYPKYNVNLNNEETIIOCINNNG 120  
 |||||||  
 DB 62 glysskspsevnisqndndiyngrynsisfwripkyfnkvnlnneytldidinnmsg 121  
 |||||||

QY 121 WKISLNNKIIITWLOTAGNCKLYEFTOMISISDYINKWIEFTITNNRNGSRYYNG 180  
 |||||||  
 DB 122 WKISLNNKIIITWLOTAGNCKLYEFTOMISISDYINKWIEFTITNNRNGSRYYNG 181  
 |||||||

QY 181 NIDKSTISNLGDIHVSNDILFKIVGCDNDRYVGIKRFKVPDELGKTEIEFLYSDEPDP 240  
 |||||||  
 DB 182 nlidkstslnlgdihvsndilfkivgcdndryvgyikrfvtdelgkteleltsdepdp 241  
 |||||||

QY 241 SILKDFMGNYLLYNKRYLLNLRTDKSITONSFLININQORGYOKPNIIFSRTLYYGV 300  
 |||||||  
 DB 242 silkdffwgnyllynkryllnlrtckslqnsflinngqygykpnlfnsrlylgy 301  
 |||||||

QY 301 EYIIRKNGSTDISNTDNFVRKNDLAYINWDRDVEYRLYADISIAKPEKIKLIRTSNSN 360  
 |||||||  
 DB 302 ewlirkgstdisntdnfvrkndlayinwdrdveyrlyadisiakpekiklirftrnsn 361  
 |||||||

QY 361 NSIGQIIVMDISIGNCTMNFQNNNGNIGLGFHSNMLYASSWYNNIRKNTSSNGCFWS 420  
 |||||||  
 DB 362 ns1gqilvmdsigmctmfnngngnig1lghfshnlvasewymnrkntsgcfs 421  
 |||||||

QY 421 FISKHEGMOEN 431  
 |||||||  
 DB 422 fiskhegwgen 432

RESULT 4  
 AAE07894  
 ID AAE07894 standard; Protein: 645 AA.  
 AC AAE07894;  
 XX  
 XX 01-NOV-2001 (first entry)  
 DE Modified clostridial heavy chain fragment #1.  
 XX  
 XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
 KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
 KW diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.  
 XX  
 XX Chimeric - Corynebacterium diphtheriae.  
 OS Chimeric - Clostridium botulinum.  
 OS  
 XX WO200158936-A2.  
 PN 16-AUG-2001.  
 XX  
 XX 04-DEC-2000; 2000MO-GB04644.  
 PF  
 XX 02-DEC-1999; 99GB-0028530.  
 PR 07-APR-2000; 2000GB-0008658.  
 PR  
 XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 PA  
 XX Shone CC, Sutton JM, Silman N;  
 PI  
 XX WPI: 2001-514643/56.  
 DR  
 XX  
 PT New non toxic polypeptide for delivery of a therapeutic agent for the  
 PT treatment of a CNS disorder comprising a binding domain that  
 PT translocates the therapeutic agent into the neuronal cells -  
 XX  
 XX Example 2; Page 44; 50pp; English.  
 XX

CC The invention relates to a non toxic polypeptide, for delivery of a  
 CC therapeutic agent to a neuronal cell, which comprises a binding domain  
 CC (carboxy terminal half of heavy chain (HC)) of a neurotoxin, designated  
 CC as HC1 that binds to the neuronal cell and a translocation domain (amino  
 CC terminal half of HC, designated as HN), that translocates the therapeutic  
 CC agent into the neuronal cell, where the translocation domain is not a HN  
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
 CC useful for the treatment of a disease state associated with neuronal  
 CC cells. The polypeptide constructs are useful for delivering therapeutic  
 CC substances to neuronal cells. They are useful to treat disorders of the  
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
 CC and infection. They are also useful in gene therapy. The present sequence  
 CC is modified clostridial heavy chain fragment. This sequence is  
 CC constructed by fusing the binding domain of botulinum neurotoxin type F  
 CC (BoNT/F) with translocation domain of diphtheria neurotoxin.  
 XX Sequence 645 AA:

Sequence 645 AA;

Query Match	100.0%	Score 431	DB 22	Length 645
Best Local Similarity	100.0%	Pred. No. 0		
Matches 431	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	SYTNDKILILYFNKRYKXIKXONSLIDMRYENKRFIDISGYGNISINDVYLYSTRNRQF	60
Db	215	sytncklllllyfnkrykxikdnslldmryenkfildisgygnislngdvylystrngf	27
QY	61	GIYSKPESEVNIACNNDDIYNGRYONFISEFWIRIPKRYENKVLNLEYIIIDICRNNSG	12
Db	275	gyyskpevevnaqndlllyngrygnfsfsfvrilpkyfnkvnlnneyllidcirmnsg	33
QY	121	WKISLNTYKKIITWTDLTAGNNOOKLYENTOMISISDIYINKWIFVITTNRLGNSRYING	18
Db	335	wkislnytkllwclqldtagngqklyfnylqmslsdylnkwlfviltmrlgnsrying	39
QY	181	NLIDEXSISNLGDIIVASNIEFKIYGCNDTRVYGRFVPTTELGKTEIETLYEDDPP	24
Db	395	nlidexsjsnlgdihvsnllflkygndtrvygrrfkyvftelgketeiellyedpp	45
QY	241	SILKDFWGNLYLKNRYLLNLRTDKSITONSFNFLINOOQGVYQKQPIESNTRYGV	30
Db	455	silkdfwgnlyllnkrlyllnlrltdksaitqsnflninqgyvqkpnifantcrlgyv	51
QY	301	EVIIEKNSSTDISNNDNVVRKNDLAYINWVRDVEYRLYADISIAKPEKTIILRTSSN	36
Db	515	eviietkngstdisndnvtvrkndlayinvvdrdvyrlyadisiakpektilrltssn	57
QY	361	NSLGOIIVMDSIGNNCTNMFONNNGNIGLGFHSNNLIVASSWYNNIRKNTSSNGCFS	42
Db	575	nslgliivmdsignnctmfnqnnngnigllgfhsnnlivaasswyynnirkntssngcfs	63
QY	421	FISKEHGQEN 431	
Db	635	fiskehgqen 645	

## RESULT 5

ID AAE07893 standard; Protein; 685 AA.

AC AAE07893-

DT 01-NOV-2001 (first entry)

Modified clostridial heavy chain-superoxide dismutase conjugate #5.

**KW** Neuronal cell; binding domain; translocation domain; stroke; epilepsy

superoxide dismutase; SOD; botulinum neurotoxin type F; BONT/F.

US Chimeric - *Bacillus steatothermophilus*.  
 OS Chimeric - Influenza virus

OS Chimeric - Clostridium botulinum.

OS Chimeric - Synthetic.

PN W0200158936-A2.

PD 16-AUG-2001.

PF 04-DEC-2000; 2000WO-GB04644

PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Shone CC, Sutton JM, Silman N;

DR WPI; 2001-514643/56.

PT New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -

PS Example 9; Page 43; 50pp; English.

CC The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin designated  
CC as Hc) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This  
CC conjugate comprises bacterial Mn superoxide dismutase (MnSOD), from  
CC *Bacillus stearothermophilus*, linker that can be cleaved by factor Xa,  
CC translocation peptide from influenza virus and a neuronal cell-specific  
CC binding domain from botulinum neurotoxin type F (BoNT/F).

**SQ**      **Sequence**      **685 AA;**

Query Match	100.0%	Score 431;	DB 22;	Length 685;
Best Local Similarity	100.0%	Pred. NO. 0;		
Matches 431; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY	1	SYNDKILILLYENKLYKKIKKONSLILDMRYENKKEDILSGVGSNLSINGDVYITSTNNQF	60
	255	SYNDKILILLYfNkLYKKIKONSILDMRYENKkILdISyGSNLSIngDVyISytInqf	314
Db			
QY	61	GIYSKSPSEVIAQNNDIYGRYQNSISFPWVPIPYFKNVNLNNEVTTIIDCIRNNNSG	120
	315	gIySkSpSeVnIaQnDILyngRyqNstIsfWvPIpPyfNkVnLnNeYlIdCIRnnnsg	374
Db			
QY	121	WKILANTNKIITWTDOPDAGNNOQKLVFVNTOMISISDYINKKIEPVTINRNLGNSRIING	180
Db	375	wKsInpNkIiWtIdqDAGnNqKlVfVntOmIsIsDyInKkIePvtIcnnlGnsrIyIng	434
QY	181	NLIDEXSISNLGDILHVSNDILFKIIVGCONDTRYVGIREFKVEDTELKTEIETLYSDEDP	240
Db	435	nLIdExsIsnLgDILhVsNdILfKlIvGcNDtRyVgIrEfKvEdTeLkTeIeTlySdeDp	494
QY	241	SILKDEMGENVLLYKKRYLLNLLEFTDKSITQNSNFMILINOORGYYQKPNIFSNTRLYTG	300
Db	495	sIlKdEgNvYlLYkKrYlLlNlLEfTdkSiTqNsNfMILInOoRGyYqKpNIfSntRlYtG	554
QY	301	EVTIRKAGSDIDSTDNFVKRNDLAIYVNDVDEYERLLYADISIAKEKIKILIRTSNSN	360
Db	555	eVtIrKAgSdIdStDnFvKrnDlAiyVnDvDeYERlLYAdISIAkEkIKILIRtSnsN	614

```
OY 361 NSLGOIIVMDSIGNNCTNMFONNNGNIGLGFHSNNLVASSWYNNIRKNTSSNGCFWS 420
DB 615 nslygqilvmdsigncctmfgnnngniglgfhsnlvasswyynnirkntssngcftws 674
OY 421 FISKEHQMOEN 431
DB 675 fiskehqmqen 685

RESULT 6
AAE07890
ID AAE07890 standard; Protein: 862 AA.
XX
AC AAE07890:
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin;
KW botulinum neurotoxin type F; BONT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
XX WO200158936-A2.
XX
XX 16-AUG-2001.
XX
XX 04-DEC-2000; 2000WO-GB04644.
XX
XX 02-DEC-1999; 99GB-0028530.
XX 07-APR-2000; 2000GB-0008658.
XX
XX (MCCR-) MICROBIOLOGICAL RES AUTHORITY.
XX
XX Shone CC, Sutton JM, Silman N;
XX
XX WPI: 2001-514643/56.
XX
XX
XX New non toxic polypeptide for delivery of a therapeutic agent for the
XX treatment of a CNS disorder comprising a binding domain that
XX translocates the therapeutic agent into the neuronal cells -
XX
XX Example 9; Page 40; 50pp; English.
XX
XX The invention relates to a non toxic polypeptide, for delivery of a
XX therapeutic agent to a neuronal cell, which comprises a binding domain
XX (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
XX as HC) that binds to the neuronal cell and a translocation domain (amino
XX terminal half of HC, designated as HN), that translocates the therapeutic
XX agent into the neuronal cell, where the translocation domain is not a HN
XX domain of a clostridial neurotoxin and is not a fragment or derivative of
XX a HN domain of a clostridial toxin. Polypeptides of the invention are
XX useful for the treatment of a disease state associated with neuronal
XX cells. The polypeptide constructs are useful for delivering therapeutic
XX substances to neuronal cells. They are useful to treat disorders of the
XX CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
XX and infection. They are also useful in gene therapy. The present sequence
XX is modified clostridial heavy chain-superoxide dismutase conjugate.
XX CC This conjugate comprises bacterial Mr-superoxide dismutase (MnSOD) from
XX CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
XX CC translocation domain from diphtheria neurotoxin and a neuronal cell-
XX CC specific binding domain from botulinum neurotoxin type F (BONT/F).
```

```
Query Match 100.0%; Score 431; DB 22; Length 862;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTNDKILLYFNKIKYKIKDNTLMRYENKRPDISGYSNISTSGDYLYSTNRNOF 60
DB 432 syndklllyfnklykkykikdnslmryemkfidlsygsnistsngdylystnrnf 491
OY 61 GIYSKRPSEYNIAQONNDIYNGKQNFSEFWNRIPKYPFNKYNLANNEYIIDCTRNNNSG 120
DB 492 gIyskRpsEynIaQonNdIyNgKqNfSEfWNrIpKYPfNKYnLANNeYIIDCTRnnNsG 551
OY 121 WKISLANKKIWFLODTAGNNQKLVFNYYQTOMISISDYINKWIFVTITNNRLGNSRIYING 180
DB 552 wkIslAnKkIwFlOdTAgNnQkLvFnYyQtOmISISdyInKwIfvTITnnRLGnsRiYing 611
OY 181 NLIDKSIISNLADIHVSDNILEFKIVGCDNTRVVGIRFEKVPDELGKTEIETLXSDPDP 240
DB 612 nlIdKsIISnLAdIhVsdNlLEfKIVgCDnTRvVGIRfEKVPdELGkTEIEtLxSDpDP 671
OY 241 SIKDPFNGVYLLYNNRRYLLNLRTQKSTONSFLNINOQGVYOKPFIENPRLYTG 300
DB 672 sIkDpFNgVYlLYnnRRyLLnLRTqKStONSfLNINoQGVyOKpFIeNPRLyTgV 731
OY 301 EVITRRKGSFSDISNTDNFVRKNDLAYINVDREVEPLXADISIAPEKIIKLRITSN 360
DB 732 eVITRRgSfSDISnTDnFvRKNDlAYINvDReVEpLxAdISIApEKIIkLRITsn 791
OY 361 NSLGOIIVMDSIGNNCTNMFONNNGNIGLGFHSNNLVASSWYNNIRKNTSSNGCFWS 420
DB 792 nslygqilvmdsigncctmfgnnngniglgfhsnlvasswyynnirkntssngcftws 851
OY 421 FISKEHQMOEN 431
DB 852 fiskehqmqen 862

RESULT 7
AAE07892
ID AAE07892 standard; Protein: 887 AA.
XX
XX AAE07892:
XX
XX 01-NOV-2001 (first entry)
XX
XX
XX Modified clostridial heavy chain-superoxide dismutase conjugate #4.
XX
XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
XX tumour; infection; neurodegenerative disease; gene therapy; chimeric;
XX superoxide dismutase; SOD; diphtheria neurotoxin; human;
XX botulinum neurotoxin type F; BONT/F.
XX
XX Chimeric - Homo sapiens.
XX Chimeric - Bacillus stearothermophilus.
XX Chimeric - Corynebacterium diphtheriae.
XX Chimeric - Clostridium botulinum.
XX Chimeric - Synthetic.
XX
XX WO200158936-A2.
XX
XX 16-AUG-2001.
XX
XX 04-DEC-2000; 2000WO-GB04644.
XX
XX 02-DEC-1999; 99GB-0028530.
XX 07-APR-2000; 2000GB-0008658.
XX
XX (MCCR-) MICROBIOLOGICAL RES AUTHORITY.
XX
XX Shone CC, Sutton JM, Silman N;
XX
XX WPI: 2001-514643/56.
XX
```

PT New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -

XX Example 9; Page 42; 50pp: English.

CC The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as HC) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is modified clostridial heavy chain superoxide dismutase conjugate.  
CC This conjugate comprises a mitochondrial leader sequence from human  
CC Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,  
CC linker that can be cleaved by thrombin, translocation domain from  
CC diphtheria neurotoxin and a neuronal cell-specific binding domain from  
CC botulinum neurotoxin type F (BontF/F).

XX Sequence 887 AA;

Query Match 100.0%; Score 431; DB 22; Length 887;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNCKILILYFNKLLKTKDKNSIDMRKFNKFDISGYSNISNGVYIYSTRNPF 60  
DB 457 sytnckililfynkllktdknsidmrkfnkfdisgysnisngvdyistrnqf 516  
QY 61 GIYSSKPESEVNIAQONDDIYNGRYONFISFWYRIKRYENKVLNNEYITIDCIRNNSG 120  
DB 517 gIySSKpSeVnIaQonDdIyNgRyOnfIsfWyrIkRyEnKvLnNeYiTIdCiRnNsG 576  
QY 121 WKSLYNNKIITWLDPTAGNOKLVENYTOISIDYINKWJFVITNNRLGNSRIYNG 180  
DB 577 wKslYnnKlIiTWldPtAGnOkLvEnYtOIsIdYInKwJfVITnnRlGnsRiYng 636  
QY 181 NLIDEKSISNLGDHNSDNLFRKIVGNCNDRYVGIRYFVETELGTEIETLYSDEPP 240  
DB 637 nLIdEkSiSnLgDhNsDnlFrKivGncNdRyVgIrYfVetElGteIeTlYsDePp 696  
QY 241 SLTKDFWGYLLYNNKRYLLNLRTDKSITONSFNLIINOQGVYOKPIFNSTRLYTGV 300  
DB 697 sLtkDfWgYlLYnnKryLLnLrTdKsItOnSfNlInOqGvYokPiFnsTrLyTgv 756  
QY 301 EYIIRKNGSTDSINDFPKNDLAIYVVDREYRLVADISIAPEKTIKIRTSNSN 360  
DB 757 eYiIrKngStDsInDfPKndLaiYvVdReYrLvAdISiApeKtIKIRtsNsN 816  
QY 361 NSLGOIIVWDSIGNCTNMFONNGNIGLGFHSNNLVAASSWYVNNIRKNTSSNGCFNS 420  
DB 817 nSlGoiIvWdsIGnCTnMfOnNgNiGlGfHsNnLvaASswYvNnIRkNtSSngCfns 876  
QY 421 FTSKEHQDEN 431  
DB 877 fTsKeHqWgen 887

RESULT 8

AAE07901 standard; Protein; 1032 AA.

AC AAE07901;  
XX  
XX 01-NOV-2001 (first entry)

XX C. botulinum C2 translocation domain with BontF/F-binding domain #2.  
DE  
XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy;  
KW botulinum neurotoxin type F; BontF/F.

OS Clostridium botulinum.

PN WO200158936-A2.

PD 16-AUG-2001.

PF 04-DEC-2000; 2000WO-GB04644.

PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Shone CC, Sutton JM, Silman N;

DR WPI: 2001-514643/56.

PT New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -

XX Example 2; Page 48; 50pp: English.

CC The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as HC) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is C. botulinum C2 enterotoxin translocation domain with botulinum  
CC neurotoxin type F (BontF/F) binding domain used in the exemplification of  
CC the invention.

XX Sequence 1032 AA;

Query Match 100.0%; Score 431; DB 22; Length 1032;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNCKILILYFNKLLKTKDKNSIDMRKFNKFDISGYSNISNGVYIYSTRNPF 60  
DB 602 sytnckililfynkllktdknsidmrkfnkfdisgysnisngvdyistrnqf 661  
QY 61 GIYSSKPESEVNIAQONDDIYNGRYONFISFWYRIKRYENKVLNNEYITIDCIRNNSG 120  
DB 662 gIySSKpSeVnIaQonDdIyNgRyOnfIsfWyrIkRyEnKvLnNeYiTIdCiRnNsG 721  
QY 121 WKSLYNNKIITWLDPTAGNOKLVENYTOISIDYINKWJFVITNNRLGNSRIYNG 180  
DB 722 wKslYnnKlIiTWldPtAGnOkLvEnYtOIsIdYInKwJfVITnnRlGnsRiYng 781  
QY 181 NLIDEKSISNLGDHNSDNLFRKIVGNCNDRYVGIRYFVETELGTEIETLYSDEPP 240  
DB 782 nLIdEkSiSnLgDhNsDnlFrKivGncNdRyVgIrYfVetElGteIeTlYsDePp 841  
QY 241 SLTKDFWGYLLYNNKRYLLNLRTDKSITONSFNLIINOQGVYOKPIFNSTRLYTGV 300  
DB 842 sLtkDfWgYlLYnnKryLLnLrTdKsItOnSfNlInOqGvYokPiFnsTrLyTgv 901

OY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINWVDRDVEYRLVADISIAKPEKIKIRTSNSN 360  
DB 902 evlirktngstdisntdnfvrkndlayinvdrdveyrlyadisiakpekikilirtsn 961  
OY 361 NSIGQIYWDSTGNKCTMNFQNNNGNIGLLGFHSNNLVASSWYNNIRKRTSNGCFMS 420  
DB 962 nsigqilywdstgncctmfnqnnngnigllgfhsnnlvasswyynnirktcsngcfms 1021  
OY 421 FISKEHMOEN 431  
DB 1022 fiskehmqen 1032

RESULT 9  
AA93309 standard; protein: 1059 AA.  
XX AAY93309;  
XX  
XX  
XX 04-SEP-2000 (first entry)  
XX  
XX A manganese superoxide dismutase (Mn-SOD) construct.  
XX  
XX Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
XX neuronal cell targeting component; NCTC; neuronal disease;  
XX oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
XX Huntington's disease; motor neurone disease;  
XX botulinum neurotoxin serotype F.  
XX  
XX Synthetic.  
XX Bacillus stearothermophilus.  
XX Clostridium botulinum.  
XX  
XX WO200028041-A1.  
XX  
XX 18-MAY-2000.  
XX  
XX 05-NOV-1999; 99WO-GB03699.  
XX  
XX 05-NOV-1998; 98GB-0024282.  
XX  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
XX PI Shone CC, Sutton JM, Hallis B, Silman N;  
XX  
XX WPI; 2000-376553/32.  
XX  
XX Novel composition, comprising superoxide dismutase linked by a  
XX cleavable linker to a neuronal cell targeting component useful for  
XX delivering superoxide dismutase to neuronal cells to treat ischemia -  
XX  
XX Disclosure: Page 48-51; 65pp; English.  
XX  
XX The present sequence represents a construct of the invention, comprising  
XX a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that  
XX can be cleaved by thrombin, and a heavy chain derived from botulinum  
XX neurotoxin serotype F. The specification describes a composition for  
XX delivery of SOD to neuronal cells. The composition comprises SOD linked,  
XX by a cleavable linker, to a neuronal cell targeting component (NCTC).  
XX This component has a domain that binds to a neuronal cell and a  
XX domain that translocates the SOD of the composition into the neuronal  
XX cell. After translocation, the linker is cleaved to release the SOD.  
XX The composition is useful for treating neuronal diseases caused or  
XX augmented by oxidative stress, such as ischemic stroke, trauma,  
XX Parkinson's disease, Huntington's disease and motor neurone diseases.  
XX  
XX Sequence 1059 AA:

Query Match 100.0%; Score 431; DB 21; Length 1059;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTNKXILLYFNKLYKKIKKDNSIIDMRENNKFIIDISGYSNISINDVYIYSTRNOF 60  
DB 629 sytnkxillyfnklykkikdksiidmremnkfidisgysnisindvyiystrnrf 688  
OY 61 GYSSKPESEVNIQANDIYNGRYONFESIFWVNIPEYRNKYNLANETIITDCIRNNNG 120  
DB 689 gyskpspevniqa ndi yngr yonfes ifwvnipeyrnkynlanetiitdcirnnng 748  
OY 121 WKISLNNKTIITWLTDTAGNNOKLVFNPTQMSISDYINKMIFVTITNNRGSRIYNG 180  
DB 749 wkisl nntiitwlt dtagnnoklvfnptqms isdyinkmifvtitnnr gsr i yng 808  
OY 181 NLIDKXSINLGDIIHWSDNILFKIVGNDTRFYVIGIRFYFVPTDELKTEITLSDPDP 240  
DB 809 nlidksisnlgdiihwsdn ilfkivgndtrfyvigirfyfvp tdelkteitl sdpdp 868  
OY 241 SILKDFWGNVLLYKRRYLLNLRTDKSITONSNFLINQORGVYOKPFIPTSLRYTGV 300  
DB 869 silkd fwnvll ykrryllnlrt dksitonsnfl inqor gvyokpfiptsl r y tgv 928  
OY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINWVDRDVEYRLVADISIAKPEKIKIRTSNSN 360  
DB 929 evlirktngstdisntdnfvrkndlayinvdrdveyrlyadisiakpekikilirtsn 988  
OY 361 NSIGQIYWDSTGNKCTMNFQNNNGNIGLLGFHSNNLVASSWYNNIRKRTSNGCFMS 420  
DB 989 nsigqilywdstgncctmfnqnnngnigllgfhsnnlvasswyynnirktcsngcfms 1048  
OY 421 FISKEHMOEN 431  
DB 1049 fiskehmqen 1059

RESULT 10  
AA93312 standard; protein: 1084 AA.  
XX AAY93312;  
XX  
XX  
XX 04-SEP-2000 (first entry)  
XX  
XX A manganese superoxide dismutase (Mn-SOD) construct.  
XX  
XX Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
XX neuronal cell targeting component; NCTC; neuronal disease;  
XX oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
XX Huntington's disease; motor neurone disease;  
XX botulinum neurotoxin serotype F.  
XX  
XX Synthetic.  
XX Homo sapiens.  
XX Bacillus stearothermophilus.  
XX Clostridium botulinum.  
XX  
XX WO200028041-A1.  
XX  
XX 18-MAY-2000.  
XX  
XX 05-NOV-1999; 99WO-GB03699.  
XX  
XX 05-NOV-1998; 98GB-0024282.  
XX  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
XX PI Shone CC, Sutton JM, Hallis B, Silman N;  
XX  
XX WPI; 2000-376553/32.  
XX  
XX Novel composition, comprising superoxide dismutase linked by a  
XX cleavable linker to a neuronal cell targeting component useful for  
XX delivering superoxide dismutase to neuronal cells to treat ischemia -

PS Disclosure: Page 57-60; 65pp; English.

XX The present sequence represents a construct of the invention, comprising

CC a mitochondrial leader sequence from human manganese superoxide

CC dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker

CC that can be cleaved by thrombin, and a heavy chain derived from

CC botulinum neurotoxin serotype F. The specification describes a

CC composition for delivery of SOD to neuronal cells. The composition

CC comprises SOD linked, by a cleavable linker, to a neuronal cell

CC targeting component (NCTC). This component has a domain that binds

CC to a neuronal cell and a domain that translocates the SOD of the

CC composition into the neuronal cell. After translocation, the linker

CC is cleaved to release the SOD. The composition is useful for treating

CC neuronal diseases caused or augmented by oxidative stress, such as

CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and

CC motor neurone diseases.

XX

Sequence 1084 AA;

SQ

Query Match 100.0%; Score 431; DB 21; Length 1084;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STTNKILILYFNKLYKKIKDNIIDMKYENKFDISGYSNISINGDVIYVYSTRNQF 60

DB 654 sytnkllillyfnklykkikdnlidmyenkfkidsgysnisngdviystrngf 713

QY 61 GYSSKPESEVNTAQNNDIYNGRYONFSEFWRIIPKYNKYNLNNEYIIDICIRNNSG 120

DB 714 gyskpeevnlaqndiyngrynfsefwripkyfinkvnlneyilidicirnnsg 773

QY 121 WKISLNYNKIITWLTODTGAONNOKLVFNQYOMISIDYINKWTFVITNNRIGNSRYING 180

DB 774 wkislnyknltwldtgaonoklvfnqymisidynkwtfvltnnrignsrying 833

QY 181 NLIDEKSISNLGDIHVSDNILEFKIVGNDTRYGVGIRFEVFPDELGKTEIETLYSDEBP 240

DB 834 nlideksisnlgdihvsnlilefkivgndtrygvgirfevfpdelgteietlysdep 893

QY 241 SILKDFWGNLYLLNKRYYLNLRLTRDKSTTONSFLNIQOAGVYOKPINFSTRLYTGV 300

DB 894 silkdfgwnlyllnkrlylnlrltrdksttonsflniqogvgyokpnifstrlytgv 953

QY 301 EYIIRKNGSTDISNTDNFVRKNDLAYINVDROVEYRLYADISIAKPEKIIKLIRTSNSN 360

DB 954 eyiirngstdisntdnfvrkndlayinvdroveyrlyadisiakpekliklirtsnsn 1013

QY 361 NSLGOIIVWDSIGNCTMNFONNNGNIGLGFHSNNLVASSWYNNIRKNTSSNGCFWS 420

DB 1014 nslgiiwdsignctmfnngngniglgfhsnnlvaasswyynnirkntssngcfws 1073

QY 421 FISKEHQEN 431

DB 1074 fiskehgwgen 1084

RESULT 11

AAE07900

ID AAE07900 standard; Protein: 1092 AA.

XX

AC AAE07900;

XX

DT 01-NOV-2001 (first entry)

XX

XX C. botulinum C2 translocation domain with BONT/F-binding domain #1.

XX

KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;

KW Tumour; infection; neurodegenerative disease; gene therapy;

KW botulinum neurotoxin type F; BONT/F.

XX

OS Clostridium botulinum.

XX

PN WO200158936-A2.

XX

XX 16-AUG-2001.

XX

PF 04-DEC-2000; 2000MO-GB04644.

XX

XX 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.

XX

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX

PI Shone CC, Sutton JM, Silman N;

XX

DR WPI; 2001-514643/56.

XX

PT New non toxic polypeptide for delivery of a therapeutic agent for the

PT treatment of a CNS disorder comprising a binding domain that

PT translocates the therapeutic agent into the neuronal cells -

XX

Example 2; Page 47; 50pp; English.

PS The invention relates to a non toxic polypeptide, for delivery of a

XX therapeutic agent to a neuronal cell, which comprises a binding domain

XX (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated

XX as HC) that binds to the neuronal cell and a translocation domain (amino

XX terminal half of HC, designated as HN), that translocates the therapeutic

XX agent into the neuronal cell, where the translocation domain is not a HN

XX domain of a clostridial neurotoxin and is not a fragment or derivative of

XX a HN domain of a clostridial toxin. Polypeptides of the invention are

XX useful for the treatment of a disease state associated with neuronal

XX cells. The polypeptide constructs are useful for delivering therapeutic

XX substances to neuronal cells. They are useful to treat disorders of the

XX CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours

XX and infection. They are also useful in gene therapy. The present sequence

XX is C. botulinum C2 enterotoxin translocation domain with botulinum

XX neurotoxin type F (BONT/F) binding domain used in the exemplification of

XX the invention.

SQ Sequence 1092 AA;

Query Match 100.0%; Score 431; DB 22; Length 1092;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STTNKILILYFNKLYKKIKDNIIDMKYENKFDISGYSNISINGDVIYVYSTRNQF 60

DB 662 sytnkllillyfnklykkikdnlidmyenkfkidsgysnisngdviystrngf 721

QY 61 GYSSKPESEVNTAQNNDIYNGRYONFSEFWRIIPKYNKYNLNNEYIIDICIRNNSG 120

DB 722 gyskpeevnlaqndiyngrynfsefwripkyfinkvnlneyilidicirnnsg 781

QY 121 WKISLNYNKIITWLTODTGAONNOKLVFNQYOMISIDYINKWTFVITNNRIGNSRYING 180

DB 782 wkislnyknltwldtgaonoklvfnqymisidynkwtfvltnnrignsrying 841

QY 181 NLIDEKSISNLGDIHVSDNILEFKIVGNDTRYGVGIRFEVFPDELGKTEIETLYSDEBP 240

DB 842 nlideksisnlgdihvsnlilefkivgndtrygvgirfevfpdelgteietlysdep 901

QY 241 SILKDFWGNLYLLNKRYYLNLRLTRDKSTTONSFLNIQOAGVYOKPINFSTRLYTGV 300

DB 902 silkdfgwnlyllnkrlylnlrltrdksttonsflniqogvgyokpnifstrlytgv 961

QY 301 EYIIRKNGSTDISNTDNFVRKNDLAYINVDROVEYRLYADISIAKPEKIIKLIRTSNSN 360

DB 962 eyiirngstdisntdnfvrkndlayinvdroveyrlyadisiakpekliklirtsnsn 1021

QY 361 NSLGOIIVWDSIGNCTMNFONNNGNIGLGFHSNNLVASSWYNNIRKNTSSNGCFWS 420

DB 1022 nslgiiwdsignctmfnngngniglgfhsnnlvaasswyynnirkntssngcfws 1081



```

QY      421 FISKHEGMOEN 431
      |||||
Db      1082 fiskhegwm 1092

RESULT 12
AAE07898
ID      AAE07898 standard; Protein; 660 AA.
AC
XX      AAE07898;
DT      01-NOV-2001 (first entry)
DE      Modified clostridial heavy chain fragment #5.
KM      Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KM      tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KM      diphtheria neurotoxin; tetanus neurotoxin; TeNT;
KM      botulinum neurotoxin type F; BONT/F.
XX      Chimeric - Corynebacterium diphtheriae.
OS      Chimeric - Clostridium tetani.
XX      Chimeric - Clostridium botulinum.
XX      MO200158936-AZ.
XX      16-AUG-2001.
XX      04-DEC-2000; 2000WO-GR04644.
XX      02-DEC-1999; 99GB-0028530.
PR      07-APR-2000; 2000GB-0008658.
XX
PA      (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PI      Shone CC, Sutton JM, Silman N;
DR      WPI: 2001-514643/56.
XX
PT      New non toxic polypeptide for delivery of a therapeutic agent for the
PT      treatment of a CNS disorder comprising a binding domain that
PT      translocates the therapeutic agent into the neuronal cells -
XX
PS      Example 2; Page 46; 50pp; English.
XX
CC      The invention relates to a non toxic polypeptide, for delivery of a
CC      therapeutic agent to a neuronal cell, which comprises a binding domain
CC      (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC      as HC) that binds to the neuronal cell and a translocation domain (amino
CC      terminal half of HC, designated as HN), that translocates the therapeutic
CC      agent into the neuronal cell, where the translocation domain is not a HN
CC      domain of a clostridial neurotoxin and is not a fragment or derivative of
CC      a HN domain of a clostridial toxin. Polypeptides of the invention are
CC      useful for the treatment of a disease state associated with neuronal
CC      cells. The polypeptide constructs are useful for delivering therapeutic
CC      substances to neuronal cells. They are useful to treat disorders of the
CC      CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC      and infection. They are also useful in gene therapy. The present sequence
CC      is modified clostridial heavy chain fragment. This sequence is
CC      constructed by fusing the binding domain which is a hybrid of botulinum
CC      neurotoxin type F (BONT/F) and tetanus neurotoxin (TeNT) domain II with
CC      translocation domain of diphtheria neurotoxin.
XX
XX      Sequence 660 AA;
XX
Query Match      49.4%; Score 213; DB 22; Length 660;
Best Local Similarity 100.0%; Pred. No. 5,7e-20;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 SYNDKILIVFNKLYKKIDNSILDMRYENKRPDISGSGSISINSGVYIYSNRP 60
      |||||
Db      215 syndklllyfnklykkidsilmsymenkfidsgysinsngvlyysnrrgt 274

```

```

QY      61 GIVSKRPEVNIAONNDIYNGRYONFSISFWRIIPKYENKVNUNNEYTIDICIRNNSSG 120
      |||||
Db      275 givskrpsevnlaqndillyngrygnfsisfwiripkyfnkvnlnneytldicirnnssg 334

QY      121 WKISLANKIWTIDDTAGNNQKLYFVTCMKISIDYINKKIIFPTTINRNLGNSRYING 180
      |||||
Db      335 wkislankliwtidqtagnnqklyfvtcamlsidylunkwlvclmnlgnstrlyng 394

QY      181 NLDEKSISMLGDHVSNDILFKIVGCDNRRYV 213
      |||||
Db      395 nldeksismlgdhvsndilfkivgcdnrryv 427

RESULT 13
AA77138
ID      AA77138 standard; Protein; 432 AA.
AC
XX      AA77138;
DT      08-MAY-2000 (first entry)
DE      Synthetic botulinum neurotoxin serotype F (BONTF) C-terminal fragment.
XX
XX      Botulinum neurotoxin; heavy chain; BONT; serotype F;
KM      C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KM      VEE; botulism; vaccine; diagnosis; drug screening.
XX
OS      Clostridium botulinum.
XX      Synthetic.
XX      WO200002524-A2.
XX      20-JAN-2000.
XX      09-JUL-1999; 99WO-US15570.
XX      10-JUL-1998; 98US-0092416.
PR      12-MAY-1999; 99US-013870.
XX
PA      (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PI      Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith LJ;
DR      WPI: 2000-160827/14.
DR      N-PSDB: AA287216.
XX
PT      Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT      toxin serotypes A-G, is used for inducing an immune response against
PT      botulinum -
XX
XX      Claim 27; Page -; 54pp; English.
XX
CC      The invention relates to novel vaccines that induce a protective immune
CC      response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F
CC      and G (BONTA-BONTG). The vaccine of the invention is novel recombinant
CC      DNA construct comprising a vector, and at least one nucleic acid
CC      fragment comprising a C-terminal heavy chain fragment (HC) from BONT
CC      serotypes A-G. In preferred embodiments of the invention, the vector is
CC      a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC      this vector results in the production of large amounts of a protein
CC      encoded by a sequence cloned into the replicon. The constructs are used
CC      to produce vaccines against botulism. The proteins can also be used as
CC      diagnostic tools for the diagnosis of botulism. The transformed host
CC      cells can be used to analyse the effectiveness of drugs and agents which
CC      inhibit toxin effects. The vaccine currently used against botulism is
CC      dangerous and expensive to produce, and contains formalin, which is very
CC      painful for the recipient. Also, the vaccine is incomplete, in that only
CC      5 of the 7 serotypes are represented in the formulation. The novel
CC      vaccine of overcomes these problems, as it is easily purified and
CC      available in large quantities. It is also expressed in the lymph nodes
CC      for a better immune response. Sequences AA77134-Y77139 represent
CC      synthetic BONT HC fragments used in the present invention. The DNA

```

CC encoding these sequences had been optimised for codon usage for  
CC expression in yeast. Note: This sequence is not given in the  
CC specification, but is decoded from the BONTF HC DNA sequence given on  
CC pages 45-46.

XX Sequence 432 AA:

Query Match 49.2%; Score 212; DB 21; Length 432;  
Best Local Similarity 100.0%; Pred. No. 3.8e-201;  
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 SFWRBIPKFKVFNKVLNNEETIIDCIRNNNSGKISLNTKRIIWTLODTAGNNOQLVFNYT 149  
DB 91 sfwrripkyfknvlnneetliidcirmnsqwkislnyknllwtlqdtagnnqklvfny 150  
QY 150 OMISISDYINKWIFVTITNNRNGSRIRYINGNLIDEKTSISNGDIHVSNDILFKIVGCD 209  
DB 151 qmisisdynkwilvtltlnrlngrlyngnlidexslnlgdlhvsdnllfkivgcd 210  
QY 210 TRYVGIRFEKVEDPELGTEIETLSDPEPSILKDFGMYLLYKRRYYLNLRTDKSI 269  
DB 211 tryvgirfkwtdelgtelctelysdepdpelkdfngnylllykrryllnlrtdksi 270  
QY 270 TONSNEFLINQGRGVYQKPNIFSNTRLTYGVE 301  
DB 271 tqnsnflninqrgvyqkpnifsntrltygve 302

#### RESULT 14

AAW09015  
ID AAW09015 standard; Protein; 144 AA.

AC AAW09015;

DT 31-MAR-1997 (first entry)

DE Immunogenic type F botulinum toxin polypeptide (aa848-991).

KW Botulinum toxin; neurotoxin; BoBt/F; immunogen; vaccine; botulism.

OS Clostridium botulinum type F strain Langeland.

XX W09641881-A1.

PD 27-DEC-1996.

PE 12-JUN-1996; 96WO-GB01409.

XX 12-JUN-1995; 95GB-0011909.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;

DR WPI: 1997-065467/06.

PT Immunogenic type F botulinum toxin polypeptide(s) - allows

XX recombinant vaccine prodn.

PS Claim 5; Page 17-18; 37pp; English.

CC Novel polypeptides (AAW09014-17) respectively comprise amino acids  
CC 848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a  
CC type F botulinum neurotoxin (BoNT/F). They lack the L chain and  
CC HN epitopes necessary for metalloprotease activity and toxin  
CC internalisation. They are free of botulinum toxin activity but can  
CC induce protective immunity to a type F botulinum toxin, making them  
CC useful for vaccine prodn. Recombinant polypeptides can be  
CC produced in transformed host cells, esp. as fusion proteins, e.g.  
CC with maltose binding protein to facilitate purification.

XX Sequence 144 AA:

Query Match 33.4%; Score 144; DB 18; Length 144;  
Best Local Similarity 100.0%; Pred. No. 2.7e-134;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFFIDISGYSNISINDGVYISTNRNQF 60  
DB 1 sytdnkililylfnklykkikdnsildmryennkfkdigsygsnsindgvystnrnqf 60  
QY 61 GYSSKPESEVNIQAONNDIITNGRVRNFSISFWRIRPKFKNVNLNETTIIDCIRNNNSG 120  
DB 61 gyssekpeevniagqndiitngrygnfsisfwrirpkfknvnlneytliidcirmnsg 120  
QY 121 WKISLNTKRIIWTLODTAGNNOQL 144  
DB 121 wkislnyknllwtlqdtagnnqkl 144

#### RESULT 15

AAW09016  
ID AAW09016 standard; Protein; 144 AA.

AC AAW09016;

DT 31-MAR-1997 (first entry)

DE Immunogenic type F botulinum toxin polypeptide (aa992-1135).

KW Botulinum toxin; neurotoxin; BoBt/F; immunogen; vaccine; botulism.

OS Clostridium botulinum type F strain Langeland.

XX W09641881-A1.

PD 27-DEC-1996.

PE 12-JUN-1996; 96WO-GB01409.

XX 12-JUN-1995; 95GB-0011909.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;

DR WPI: 1997-065467/06.

PT Immunogenic type F botulinum toxin polypeptide(s) - allows

XX recombinant vaccine prodn.

PS Claim 5; Page 18-19; 37pp; English.

CC Novel polypeptides (AAW09014-17) respectively comprise amino acids  
CC 848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a  
CC type F botulinum neurotoxin (BoNT/F). They lack the L chain and  
CC HN epitopes necessary for metalloprotease activity and toxin  
CC internalisation. They are free of botulinum toxin activity but can  
CC induce protective immunity to a type F botulinum toxin, making them  
CC useful for vaccine prodn. Recombinant polypeptides can be  
CC produced in transformed host cells, esp. as fusion proteins, e.g.  
CC with maltose binding protein to facilitate purification.

XX Sequence 144 AA:

Query Match 33.4%; Score 144; DB 18; Length 144;  
Best Local Similarity 100.0%; Pred. No. 2.7e-134;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 VFNKYOMISISDYINKWIFVTITNNRNGSRIRYINGNLIDEKTSISNGDIHVSNDILFKI 204  
DB 1 vfnygmisisdynkwilvtltlnrlngrlyngnlidexslnlgdlhvsdnllfkil 60

Thu Aug 15 12:38:14 2002

us-08-981-087a-1.rag

Page 11

Qy 205 VANDTFRVGVGRKPKFDFELGKREFTLISDSDPSITLKQFMCNTLYLNKRYVLTNLR 264

Dd 61 VQNDTRYVGRYFRKFDCEIGKCEIELVSDDEPDSITKQFMCNTLYLNKRYVLTNLR 120

Qy 265 TKSSTQSNFNLINNOORVQKRP 288

Dd 121 LKSLTQSNFNLINNOGRVYQKRP 144

Search completed: August 15, 2002, 11:12:25  
Job time: 317 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2002, 09:39:06 ; Search time 96.3 Seconds

(without alignments)  
497.122 Million cell updates/sec

Title: US-08-981-087a-1

Perfect score: 2288  
Sequence: 1 SYTNKILILYFNKLYKKIK.....TSNCGWFSKHEGMOEN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_032802:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2288	100.0	431	18	AAW09014
2	2288	100.0	432	22	AAW04096
3	2288	100.0	432	22	AAW04103
4	2288	100.0	645	22	AAE07894
5	2288	100.0	685	22	AAE07893
6	2288	100.0	862	22	AAE07890
7	2288	100.0	887	22	AAE07892
8	2288	100.0	1032	22	AAE07901
9	2288	100.0	1059	21	AAV93309
10	2288	100.0	1084	21	AAV93312
11	2288	100.0	1092	22	AAE07900

12	2271	99.3	432	21	AAV77138
13	1800	76.7	448	19	AAW68399
14	1451.5	63.4	449	21	AAV77137
15	1451.5	63.4	449	22	AAW04094
16	1447.5	63.3	452	19	AAW68396
17	1426	62.3	451	19	AAW68395
18	1423.5	62.2	419	22	AAW04095
19	1355.5	59.2	419	22	AAE07898
20	1120.5	49.0	1067	21	AAV93307
21	1120.5	49.0	1092	21	AAV93310
22	1110.5	48.7	1296	17	AAW04088
23	1113.5	48.7	437	22	AAW04088
24	1113.5	48.7	438	19	AAW68399
25	1113.5	48.7	438	19	AAW68399
26	1113.5	48.7	438	21	AAV77134
27	1113.5	48.7	445	19	AAW68391
28	1113.5	48.7	462	17	AAW68390
29	1113.5	48.7	462	19	AAW68390
30	1111.5	48.6	434	22	AAW04089
31	1111.5	48.6	435	22	AAW04090
32	1106.5	48.4	837	21	AAV77142
33	1099.5	48.1	432	21	AAV77142
34	1088.5	47.6	847	22	AAW04081
35	1079.5	47.6	415	22	AAW04083
36	1019.5	44.6	382	21	AAW68393
37	772	33.7	472	19	AAW68393
38	770	33.7	144	18	AAW09015
39	769	33.6	1070	21	AAV93308
40	769	33.6	1095	21	AAV93311
41	769	33.6	1291	19	AAW68392
42	761	33.3	144	18	AAW09016
43	757	33.1	143	18	AAW09017
44	751	32.8	472	19	AAW68394
45	743	32.5	848	22	AAW04082

#### ALIGNMENTS

RESULT 1	
ID	AAW09014 standard; Protein; 431 AA.
AAW09014	
AC	AAW09014:
XX	
DT	31-MAR-1997 (first entry)
XX	
DE	Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX	
KW	Botulinum toxin; neurotoxin; BoNT/F; immunogen; vaccine; botulism.
XX	
OS	Clostridium botulinum type F strain Langeland.
XX	
PN	W09641881-A1.
XX	
PD	27-DEC-1996.
XX	
PF	12-JUN-1996; 96W0-GB01409.
XX	
PR	12-JUN-1995; 95GB-0011909.
XX	
PA	(MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX	
PI	Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX	
DR	WPI: 1997-065467/06.
XX	
PT	N-PSDB; AAT48100.
XX	
PS	Immunogenic type F botulinum toxin polypeptide(s) - allows
XX	
PS	recombinant vaccine prodn.
XX	
PS	Claim 5; Page 16-17; 37pp; English.



Db	362	ns1gqilwmslgnctnmfqnngnig11gfhsmnlvawswynn1rktnsngcfs	421
Oy	421	FISKEHGOEN	431
Db	422	fiskehwgen	432
	RESULT	3	
	AAAB04103		
AC	AAAB04103	standard; Protein: 432 AA.	
XX	AAAB04103;		
DT	11-APR-2001	(first entry)	
DE	Botulinum toxin heavy chain C-terminal sequence (serotype F).		
FW	Botulin; toxin; neurotoxin; heavy chain; recombinant expression; recombinant vector; antigen; immune response; vaccine; bacterium; infection.		
XX			
OS	Synthetic.		
PN	Clostridium botulinum.		
XX	WC020067700-A2.		
PD	16-NOV-2000.		
PF	12-MAY-2000; 2000MO-US12890.		
PR	12-MAY-1999; 99US-0133865.		
PR	12-MAY-1999; 99US-0133866.		
PR	12-MAY-1999; 99US-0133867.		
PR	12-MAY-1999; 99US-0133868.		
PR	12-MAY-1999; 99US-0133869.		
PR	12-MAY-1999; 99US-0133873.		
PR	29-JUL-1999; 99US-0146192.		
XX			
XX	(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.		
FA			
FI	Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;		
XX			
DR	WPI: 2001-016048/02.		
DR	N-PSDB; AAA54499.		
PT	New nucleic acids encoding the carboxy- or amino-terminal portions of the heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine against botulinism		
XX			
PS	Disclosure: Fig 18b; 73pp: English.		
CC	Botulinum neurotoxins are translated as a single 150 kDa polypeptide chain and then posttranslationally nicked, forming a dication consisting of a 100 kDa heavy chain and a 50 kDa light chain which remain linked by a disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT) can be used in recombinant expression vectors and expressed in transformed cells to produce protective immunity against botulinum neurotoxin, which causes botulism. The nucleic acids are expressible in a recombinant organism such as <i>Escherichia coli</i> or <i>Pleura pastoris</i> . The use of recombinant nucleic acids are advantageous since it eliminates the need to culture large quantities of hazardous toxin-producing bacterium. Production yield from the genetically engineered product is also high and cost of production is lower. The nucleic acids can be derived from Clostridium botulinum serotypes A-G.		
XX			
XX	Sequence	432 AA;	
XX			
Query Match	100.0%;	Score 2288;	DB 22; Length 432;
Best Local Similarity	100.0%;	Pred. NO. 8e-168;	

Matches	4331: Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
Oy	1	SYTNDKLLILYFNFKYKKIKNDSTLDMREKNNKFTLDSGYGNSINISNDVIYISNRRNF	60					
Dd	2	sytrndkllillyfnfklykkikndslldmryennkfidsygsynslsngvlystnrrnf	61					
Oy	61	GYSSKSESVNTAONNDIYNGXRYONESIFWVGIPEKYPKYNLNEETIITDTRNNSG	120					
Dd	62	gyssksepsevnadqmdllyngvyrqntslstvwrlfpyrlnkynlneetylldcltrnns	121					
Oy	121	WKISLNTNKTIWTLDTAGNNQKVLNVTOMISISDYINKKIFVTITNNRLGNSRIYNG	180					
Dd	122	wklslnyvkliwldqetagnmqkrlvtnyqmaslstdylnkwlflvtltnnrlgnsrilyng	181					
Oy	181	NLIDEKSIASMGDIHVSDNLUEFKIVSCNFTRYVGIIRFVKVPTDELGTETLTYSDEDP	240					
Dd	182	nlldeksislmgdlhvsdnllfklvgcndtryvlyrfkvlftelgtelgtelysdepp	241					
Oy	241	SLTKDFGNLILYFNKRYLNLNLRTKSTITONSFEILNINOQGVQKPNFSWTRLTYGV	300					
Dd	242	sltkdfgnlylfnkryyllnlrltrkstsitonsfelnlnqgvqkpnfswltrlygv	301					
Oy	301	EVIRKNGSFDISNTDNFVRKNDLATINVDDEYRYLADISIAKPEKITYKILFRSNSN	360					
Dd	302	evlrkngsfdisntdnfvrkndlaylnvdrdveyrlyadislakpekylkllrtsn	361					
Oy	361	NSLGQIITVMOSTGNCTMKNPNNNGNIGLGFHSNNLVASSMYNNIRKNTSSNCFFMS	420					
Dd	362	nslgqilvmstgnctmknnpnnngnlglgfhsnnlvassmynnlrkntssngcflms	421					
Oy	421	FTSKRHQMOEN	431					
Dd	422	flskrhqmoen	432					
RESULT	4							
AAE07894								
ID	AAE07894	standard; Protein: 645	AA.					
XX								
AC	AAE07894;							
XX								
DT	01-NOV-2001	(first entry)						
DE								
XX		Modified clostridial heavy chain fragment #1.						
KW		Neuronal cell; binding domain; translocation domain; stroke; epilepsy;						
KM		tumour; infection; neurodegenerative disease; gene therapy; chimeric;						
KW		diphtheria neurotoxin; botulinum neurotoxin type F; BONT/F.						
XX								
OS		Chimeric - Corynebacterium diphtheriae.						
OS		Chimeric - Clostridium botulinum.						
XX								
PN	WC0200158936-A2.							
XX								
XX	16-AUG-2001.							
XX								
PF	04-DEC-2000; 2000MO-GB04644.							
XX								
FR	02-DEC-1999; 99GB-0028550.							
FR	07-APR-2000; 2000GB-0008658.							
XX								
FA	(MICR-) MICROBIOLOGICAL RES AUTHORITY.							
XX								
PI	Shone CC, Sutton JM, Silman N;							
XX								
DR	WPI: 2001-514643/56.							
XX								
PT	New non toxic polypeptide for delivery of a therapeutic agent for the							
PT	treatment of a CNS disorder comprising a binding domain that							
PT	translocates the therapeutic agent into the neuronal cells -							
XX								
XX	Example 2; Page 44; 50pp; English.							
XX								

CC The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as HC) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is modified clostridial heavy chain fragment. This sequence is  
CC constructed by fusing the binding domain of botulinum neurotoxin type F  
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.  
XX  
SQ Sequence 645 AA;

Query Match 100.0%; Score 2288; DB 22; Length 645;  
Best Local Similarity 100.0%; Pred. No. 1.3e-167;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNDKILILYFNKLYKKIKDINSILDMRYENKFFIDISGYGNSISINGDYIYSTNRNOP 60  
DB 215 SYTNDKILILYFNKLYKKIKDINSILDMRYENKFFIDISGYGNSISINGDYIYSTNRNQF 274  
QY 61 GIYSKRPSEVNIAQNDIYNGRYONFISFWRIKRYEKVNLNNEYTTIIDCIRNNSG 120  
DB 275 GIYSKRPSEVNIAQNDIYNGRYONFISFWRIKRYEKVNLNNEYTTIIDCIRNNSG 334  
QY 121 WKISLWYKTIWLTOTAGNOKLVFNRYTOMISIDYINKMIFVTITNRLGSRITYNG 180  
DB 335 WKISLWYKTIWLTOTAGNOKLVFNRYTOMISIDYINKMIFVTITNRLGSRITYNG 334  
QY 181 NLIDESISNLGDIHVSNDILFKIVGCDTRYGIRYKRVFDELKTELETYSDEPDP 240  
DB 395 NLIDESISNLGDIHVSNDILFKIVGCDTRYGIRYKRVFDELKTELETYSDEPDP 454  
QY 241 SILKDFWGNVLLYKRYLLNLRTDKSTIONSFLINQOQGVOKPNEFSMRLYTG 300  
DB 455 SILKDFWGNVLLYKRYLLNLRTDKSTIONSFLINQOQGVOKPNEFSMRLYTG 514  
QY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINVDREVRLYADISIAKPEKIKILRTSSNS 360  
DB 515 EVIIRKNGSTDISNTDNFVRKNDLAYINVDREVRLYADISIAKPEKIKILRTSSNS 574  
QY 361 NSIGQIIVWDSIGNCTMNFQNNNGNIGLGFHSNNNLVASSWYNNIRKNTSSNGCFWS 420  
DB 575 NSIGQIIVWDSIGNCTMNFQNNNGNIGLGFHSNNNLVASSWYNNIRKNTSSNGCFWS 634  
QY 421 FISEKGMQEN 431  
DB 635 FISEKGMQEN 645

RESULT 5  
AAE07893  
ID AAE07893 standard; Protein: 685 AA.  
AC AAE07893;  
XX  
XX 01-NOV-2001 (first entry)  
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.  
XX  
XX  
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.  
XX  
OS Chimeric - Bacillus steartothermophilus.  
OS Chimeric - Influenza virus.

OS Chimeric - Clostridium botulinum.  
OS Chimeric - Synthetic.  
XX  
XX MO200158936-A2.  
XX  
XX 16-AUG-2001.  
XX  
XX 04-DEC-2000; 2000WO-GB04644.  
XX  
XX 02-DEC-1999; 99GB-0028530.  
XX 07-APR-2000; 2000GB-0008658.  
XX  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
XX Shone CC, Sutton JM, Silman N;  
XX  
XX WPI; 2001-514643/56.  
XX  
XX  
XX New non toxic polypeptide for delivery of a therapeutic agent for the  
XX treatment of a CNS disorder comprising a binding domain that  
XX translocates the therapeutic agent into the neuronal cells -  
PS Example 9; Page 43; 50pp; English.

CC The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as HC) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This  
CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from  
CC Bacillus steartothermophilus, linker that can be cleaved by factor Xa,  
CC translocation peptide from influenza virus and a neuronal cell-specific  
CC binding domain from botulinum neurotoxin type F (BoNT/F).  
XX  
SQ Sequence 685 AA;

Query Match 100.0%; Score 2288; DB 22; Length 685;  
Best Local Similarity 100.0%; Pred. No. 1.4e-167;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNDKILILYFNKLYKKIKDINSILDMRYENKFFIDISGYGNSISINGDYIYSTNRNOP 60  
DB 255 SYTNDKILILYFNKLYKKIKDINSILDMRYENKFFIDISGYGNSISINGDYIYSTNRNQF 314  
QY 61 GIYSKRPSEVNIAQNDIYNGRYONFISFWRIKRYEKVNLNNEYTTIIDCIRNNSG 120  
DB 315 GIYSKRPSEVNIAQNDIYNGRYONFISFWRIKRYEKVNLNNEYTTIIDCIRNNSG 374  
QY 121 WKISLWYKTIWLTOTAGNOKLVFNRYTOMISIDYINKMIFVTITNRLGSRITYNG 180  
DB 375 WKISLWYKTIWLTOTAGNOKLVFNRYTOMISIDYINKMIFVTITNRLGSRITYNG 434  
QY 181 NLIDESISNLGDIHVSNDILFKIVGCDTRYGIRYKRVFDELKTELETYSDEPDP 240  
DB 435 NLIDESISNLGDIHVSNDILFKIVGCDTRYGIRYKRVFDELKTELETYSDEPDP 494  
QY 241 SILKDFWGNVLLYKRYLLNLRTDKSTIONSFLINQOQGVOKPNEFSMRLYTG 300  
DB 495 SILKDFWGNVLLYKRYLLNLRTDKSTIONSFLINQOQGVOKPNEFSMRLYTG 554  
QY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINVDREVRLYADISIAKPEKIKILRTSSNS 360  
DB 555 EVIIRKNGSTDISNTDNFVRKNDLAYINVDREVRLYADISIAKPEKIKILRTSSNS 614





PT New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -

PS Example 9; Page 42; 50pp; English.

XX The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as HC) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is modified clostridial heavy chain-superoxide dismutase conjugate.  
CC This conjugate comprises a mitochondrial leader sequence from human  
CC Mn-superoxide dismutase (MnSOD). MnSOD from Bacillus stearothermophilus,  
CC linker that can be cleaved by thrombin, translocation domain from  
CC diptheria neurotoxin and a neuronal cell-specific binding domain from  
CC botulinum neurotoxin type F (BoNT/F).

CC Sequence 887 AA;

CC Sequence 887 AA;

Query Match 100.0%; Score 2288; DB 22; Length 887;  
Best Local Similarity 100.0%; Pred. No. 2e-167;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYNDKILILYFNKLYKKIKDNLDMRYENKFIIDISGYSNLSINGDVIYSTNRNOF 60  
DB 457 SYNDKILILYFNKLYKKIKDNLDMRYENKFIIDISGYSNLSINGDVIYSTNRNGF 516  
QY 61 GIYSSKRPSEVNIAQNDIYNGRYONFSIFWVRIRPKYFKVNLNNEYTTIIDCIRNNNSG 120  
DB 517 GIYSSKRPSEVNIAQNDIYNGRYONFSIFWVRIRPKYFKVNLNNEYTTIIDCIRNNNSG 576  
QY 121 WKISLVNKRKIWTLOPAGNKKLVFNRYTOMISIDYINKWIEVTTNNRGLNSRIYNG 180  
DB 577 WKISLVNKRKIWTLOPAGNKKLVFNRYTOMISIDYINKWIEVTTNNRGLNSRIYNG 636  
QY 181 NLDEKSIISNLGDIHNSDNLFKIVGCDTRRYGIRYFKVDELEKTELETLYSDDEPP 240  
DB 637 NLDEKSIISNLGDIHNSDNLFKIVGCDTRRYGIRYFKVDELEKTELETLYSDDEPP 696  
QY 241 SILKDFWGNLILYNNKRYLLNLRTDKSITQNSNLFNLINOQRGVYOKRPNFESNRLTYGV 300  
DB 697 SILKDFWGNLILYNNKRYLLNLRTDKSITQNSNLFNLINOQRGVYOKRPNFESNRLTYGV 756  
QY 301 EVIIRKNGSDISNTDNFVAKNDLAYINVDREVERLYADISIAKEKTIKLRISNSN 360  
DB 757 EVIIRKNGSDISNTDNFVAKNDLAYINVDREVERLYADISIAKEKTIKLRISNSN 816  
QY 361 NSIGQIIVMDSIGNNCTMNFQNNNGNIGLGFHSNNLVASSVYNNIRKNTSNGCEWS 420  
DB 817 NSIGQIIVMDSIGNNCTMNFQNNNGNIGLGFHSNNLVASSVYNNIRKNTSNGCEWS 876  
QY 421 FISKEHGMDEN 431  
DB 877 FISKEHGMDEN 887

RESULT 8  
AAE07901  
ID AAE07901 standard; Protein; 1032 AA.  
XX  
AC AAE07901;  
XX  
DT 01-NOV-2001 (first entry)

XX C. botulinum C2 translocation domain with BoNT/F-binding domain #2.  
DE  
XX  
XX Neuronal cell: binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy;  
KW botulinum neurotoxin type F; BoNT/F.

OS Clostridium botulinum.

PN WO200158936-A2.

PD 16-AUG-2001.

PF 04-DEC-2000; 2000WO-GB04644.

PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Shone CC, Sutton JM, Silman N;

DR WPI; 2001-514643/56.

PT New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -

PS Example 2; Page 48; 50pp; English.

XX The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as HC) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is C. botulinum C2 enterotoxin translocation domain with botulinum  
CC neurotoxin type F (BoNT/F) binding domain used in the exemplification of  
CC the invention.

CC Sequence 1032 AA;

CC Sequence 1032 AA;

Query Match 100.0%; Score 2288; DB 22; Length 1032;  
Best Local Similarity 100.0%; Pred. No. 2.4e-167;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYNDKILILYFNKLYKKIKDNLDMRYENKFIIDISGYSNLSINGDVIYSTNRNOF 60  
DB 602 SYNDKILILYFNKLYKKIKDNLDMRYENKFIIDISGYSNLSINGDVIYSTNRNGF 661  
QY 61 GIYSSKRPSEVNIAQNDIYNGRYONFSIFWVRIRPKYFKVNLNNEYTTIIDCIRNNNSG 120  
DB 662 GIYSSKRPSEVNIAQNDIYNGRYONFSIFWVRIRPKYFKVNLNNEYTTIIDCIRNNNSG 721  
QY 121 WKISLVNKRKIWTLOPAGNKKLVFNRYTOMISIDYINKWIEVTTNNRGLNSRIYNG 180  
DB 722 WKISLVNKRKIWTLOPAGNKKLVFNRYTOMISIDYINKWIEVTTNNRGLNSRIYNG 781  
QY 181 NLDEKSIISNLGDIHNSDNLFKIVGCDTRRYGIRYFKVDELEKTELETLYSDDEPP 240  
DB 782 NLDEKSIISNLGDIHNSDNLFKIVGCDTRRYGIRYFKVDELEKTELETLYSDDEPP 841  
QY 241 SILKDFWGNLILYNNKRYLLNLRTDKSITQNSNLFNLINOQRGVYOKRPNFESNRLTYGV 300  
DB 842 SILKDFWGNLILYNNKRYLLNLRTDKSITQNSNLFNLINOQRGVYOKRPNFESNRLTYGV 901

OY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDREYRLYADISIAKEPKIKILRTSN 360  
DB 902 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDREYRLYADISIAKEPKIKILRTSN 961  
OY 361 NSLGOITVWDSIGNNCTMNFQNNNGNIGLGFHSNNLVASWYNNIRKNTSSNGCPWS 420  
DB 962 NSLGOITVWDSIGNNCTMNFQNNNGNIGLGFHSNNLVASWYNNIRKNTSSNGCPWS 1021  
OY 421 FISKEHMOEN 431  
DB 1022 FISKEHMOEN 1032

RESULT 9  
AAV93309 standard; protein: 1059 AA.  
ID AAV93309;  
AC AAV93309;  
XX 04-SEP-2000 (first entry)  
XX A manganese superoxide dismutase (Mn-SOD) construct.  
XX DE  
XX KM Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
XX KM neuronal cell targeting component; NCTC; neuronal disease;  
XX KM oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
XX KM Huntington's disease; motor neurone disease;  
XX KM botulinum neurotoxin serotype F.  
XX OS Synthetic.  
XX OS Bacillus steaerothermophilus.  
XX OS Clostridium botulinum.  
XX PN W0200028041-A1.  
XX PD 18-MAY-2000.  
XX PF 05-NOV-1999; 99MO-GB03699.  
XX PR 05-NOV-1998; 98GB-0024282.  
XX PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX PI Shone CC, Sutton JM, Hallis B, Silman N;  
XX WPI: 2000-376553/32.  
XX DR Novel composition, comprising superoxide dismutase linked by a  
XX PT cleavable linker to a neuronal cell targeting component useful for  
XX PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
XX PS Disclosure: Page 48-51; 65pp; English.  
XX XX The present sequence represents a construct of the invention, comprising  
XX CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that  
XX CC can be cleaved by thrombin, and a heavy chain derived from botulinum  
XX CC neurotoxin serotype F. The specification describes a composition for  
XX CC delivery of SOD to neuronal cells. The composition comprises SOD linked,  
XX CC by a cleavable linker, to a neuronal cell targeting component (NCTC).  
XX CC This component has a domain that binds to a neuronal cell and a  
XX CC domain that translocates the SOD of the composition into the neuronal  
XX CC cell. After translocation, the linker is cleaved to release the SOD.  
XX CC The composition is useful for treating neuronal diseases caused or  
XX CC augmented by oxidative stress, such as ischemic stroke, trauma,  
XX CC Parkinson's disease, Huntington's disease and motor neurone diseases.  
XX SQ Sequence 1059 AA;

Query Match 100.0%; Score 2288; DB 21; Length 1059;  
Best Local Similarity 100.0%; Pred. No. 2.5e-167;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTNDKILLVFNKLYKKIRKNDSTDMRENNKFTDISGNSNISTNGVYTSNRNOF 60  
DB 629 SYTNDKILLVFNKLYKKIRKNDSTDMRENNKFTDISGNSNISTNGVYTSNRNOF 668  
OY 61 GIYSKRPSEVNTAQNNDITVYNGRYQNFSEFWRIPEKFNKVLNNETIITDCIRNNNG 120  
DB 669 GIYSKRPSEVNTAQNNDITVYNGRYQNFSEFWRIPEKFNKVLNNETIITDCIRNNNG 748  
OY 121 WKISLANKRIITWLODTAGNNOKLVFNQYOMISISDYINKWIEVYTTNRNLGNSRIYNG 180  
DB 749 WKISLANKRIITWLODTAGNNOKLVFNQYOMISISDYINKWIEVYTTNRNLGNSRIYNG 808  
OY 181 NLIDEXISNLGDIHVSNDNILEFKYVCNDTFRVGIKRYKVPDTELKTEIEITLYSDPEDP 240  
DB 809 NLIDEXISNLGDIHVSNDNILEFKYVCNDTFRVGIKRYKVPDTELKTEIEITLYSDPEDP 868  
OY 241 SILKDFWGNVLLYNKRYLLNLRTDQSTONSFLINQORGVQKPNIFSNRLYGV 300  
DB 869 SILKDFWGNVLLYNKRYLLNLRTDQSTONSFLINQORGVQKPNIFSNRLYGV 928  
OY 301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDREYRLYADISIAKEPKIKILRTSN 360  
DB 929 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDREYRLYADISIAKEPKIKILRTSN 968  
OY 361 NSLGOITVWDSIGNNCTMNFQNNNGNIGLGFHSNNLVASWYNNIRKNTSSNGCPWS 420  
DB 969 NSLGOITVWDSIGNNCTMNFQNNNGNIGLGFHSNNLVASWYNNIRKNTSSNGCPWS 1048  
OY 421 FISKEHMOEN 431  
DB 1049 FISKEHMOEN 1059

RESULT 10  
AAV93312 standard; protein: 1084 AA.  
ID AAV93312;  
AC AAV93312;  
XX 04-SEP-2000 (first entry)  
XX DE  
XX KM A manganese superoxide dismutase (Mn-SOD) construct.  
XX KM Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
XX KM neuronal cell targeting component; NCTC; neuronal disease;  
XX KM oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
XX KM Huntington's disease; motor neurone disease;  
XX KM botulinum neurotoxin serotype F.  
XX OS Synthetic.  
XX OS Homo sapiens.  
XX OS Bacillus steaerothermophilus.  
XX OS Clostridium botulinum.  
XX PN W0200028041-A1.  
XX PD 18-MAY-2000.  
XX PF 05-NOV-1999; 99MO-GB03699.  
XX PR 05-NOV-1998; 98GB-0024282.  
XX PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX PI Shone CC, Sutton JM, Hallis B, Silman N;  
XX WPI: 2000-376553/32.  
XX DR Novel composition, comprising superoxide dismutase linked by a  
XX PT cleavable linker to a neuronal cell targeting component useful for  
XX PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
XX XX

PS Disclosure; Page 57-60; 65pp; English.  
XX  
CC The present sequence represents a construct of the invention, comprising  
CC a mitochondrial leader sequence from human manganese superoxide  
CC dismutase (Mn-SOD), a *Bacillus stearothermophilus* Mn-SOD, a linker  
CC that can be cleaved by thrombin, and a heavy chain derived from  
CC botulinum neurotoxin serotype F. The specification describes a  
CC composition for delivery of SOD to neuronal cells. The composition  
CC comprises SOD linked, by a cleavable linker, to a neuronal cell  
CC targeting component (NCTC). This component has a domain that binds  
CC to a neuronal cell and a domain that translocates the SOD of the  
CC composition into the neuronal cell. After translocation, the linker  
CC is cleaved to release the SOD. The composition is useful for treating  
CC neuronal diseases caused or augmented by oxidative stress, such as  
CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and  
CC motor neurone diseases.  
XX  
SQ Sequence 1084 AA:  
  
Query Match 100.0%; Score 2288; DB 21; Length 1084;  
Best Local Similarity 100.0%; Pred. No. 2,6e-167;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYTNKXILLYFNKLYKKIKDINSILDMRYENKFPIDISGYSNISTNGDVIYISTNRNQF 60  
DB 654 syndklllyfinklykkikdinsildmryenkfidisgysnistsngdviystnrngf 713  
QY 61 GIYSSKPESEVNIQONNDIYNGRYQNFISFWIRIPKRYKVNLMNNEYTIIDCIRNNNSG 120  
DB 714 glyskspevevniqondliyngrymisfswiripkrykvnlmnneytliidcirnnns 773  
QY 121 WKISLWYKLIWTLQDTAGNNOQLVFNRYTOMISIDYINKWIEVTITNNRLGSRIRYING 180  
DB 774 wkislwnklywltqdtagnnqlvfnrytomsidsyinkwievtitnnrlgsriryng 833  
QY 181 NLIDESISMLGDIHVSNDILFEKIVGCDTRRYGIRFKFDELGETELETYSDEPDP 240  
DB 834 nlideksismldghvsndilfekivgcdtrrygiryfkfdeletetletysdepdp 893  
QY 241 SILKDFWGNLYLNKRYLLNLRTDKSTIQNSNFTLINOGRVYOKPNIESTRLTYGV 300  
DB 894 silkdftwgnlylnkryllnlrtkdstiqnsnftlinoqrvyokpniestrltygv 953  
QY 301 EVIIRKNGSTDISTNDFVAKNDLAYINVDREVRLYADISIAPEKTIKILRTSNNSN 360  
DB 954 eviirngstdistndfvrkndlayinvdrvevrylyadislapekikilrtsnnsn 1013  
QY 361 NSLGOITVMDISGNCTMNFQNNNGNIGLGHSHNNLVASSWYNNIRKNTSSNGCFWS 420  
DB 1014 nslgqitvmdisgnctmfnqnnngnigllghshnnlvasswyynnirkntssngcfws 1073  
QY 421 FISKEHMOEN 431  
DB 1074 fiskehngwen 1084  
  
RESULT 11  
AAE07900  
ID AAE07900 standard; Protein; 1092 AA.  
XX  
AC AAE07900;  
XX  
DT 01-NOV-2001 (first entry)  
XX  
DE C. botulinum C2 translocation domain with BONT/F-binding domain #1.  
XX  
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy;  
KW botulinum neurotoxin type F; BONT/F.  
XX  
OS Clostridium botulinum.  
XX

PN WO200158936-A2.  
XX  
XX 16-AUG-2001.  
XX  
PF 04-DEC-2000; 2000MO-GB04644.  
XX  
XX 02-DEC-1999; 99GB-0028530.  
PR 07-APR-2000; 2000GB-0008658.  
XX  
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
PI Shone CC, Sutton JM, Silman N;  
XX WPI: 2001-514643/56.  
XX  
XX  
XX New non toxic polypeptide for delivery of a therapeutic agent for the  
XX treatment of a CNS disorder comprising a binding domain that  
XX translocates the therapeutic agent into the neuronal cells -  
XX  
XX Example 2; Page 47; 50pp; English.

The invention relates to a non toxic polypeptide, for delivery of a  
therapeutic agent to a neuronal cell, which comprises a binding domain  
(carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
as Hc) that binds to the neuronal cell and a translocation domain (amino  
terminal half of HC, designated as HN), that translocates the therapeutic  
agent into the neuronal cell, where the translocation domain is not a HN  
domain of a clostridial neurotoxin and is not a fragment or derivative of  
a HN domain of a clostridial toxin. Polypeptides of the invention are  
useful for the treatment of a disease state associated with neuronal  
cells. The polypeptide constructs are useful for delivering therapeutic  
substances to neuronal cells. They are useful to treat disorders of the  
CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
and infection. They are also useful in gene therapy. The present sequence  
is C. botulinum C2 enterotoxin translocation domain with botulinum  
neurotoxin type F (BONT/F) binding domain used in the exemplification of  
the invention.

SQ Sequence 1092 AA:

Query Match 100.0%; Score 2288; DB 22; Length 1092;  
Best Local Similarity 100.0%; Pred. No. 2,6e-167;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYTNKXILLYFNKLYKKIKDINSILDMRYENKFPIDISGYSNISTNGDVIYISTNRNQF 60  
DB 662 syndklllyfinklykkikdinsildmryenkfidisgysnistsngdviystnrngf 721  
QY 61 GIYSSKPESEVNIQONNDIYNGRYQNFISFWIRIPKRYKVNLMNNEYTIIDCIRNNNSG 120  
DB 722 glyskspevevniqondliyngrymisfswiripkrykvnlmnneytliidcirnnns 781  
QY 121 WKISLWYKLIWTLQDTAGNNOQLVFNRYTOMISIDYINKWIEVTITNNRLGSRIRYING 180  
DB 782 wkislwnklywltqdtagnnqlvfnrytomsidsyinkwievtitnnrlgsriryng 841  
QY 181 NLIDESISMLGDIHVSNDILFEKIVGCDTRRYGIRFKFDELGETELETYSDEPDP 240  
DB 842 nlideksismldghvsndilfekivgcdtrrygiryfkfdeletetletysdepdp 901  
QY 241 SILKDFWGNLYLNKRYLLNLRTDKSTIQNSNFTLINOGRVYOKPNIESTRLTYGV 300  
DB 902 silkdftwgnlylnkryllnlrtkdstiqnsnftlinoqrvyokpniestrltygv 961  
QY 301 EVIIRKNGSTDISTNDFVAKNDLAYINVDREVRLYADISIAPEKTIKILRTSNNSN 360  
DB 962 eviirngstdistndfvrkndlayinvdrvevrylyadislapekikilrtsnnsn 1021  
QY 361 NSLGOITVMDISGNCTMNFQNNNGNIGLGHSHNNLVASSWYNNIRKNTSSNGCFWS 420  
DB 1022 nslgqitvmdisgnctmfnqnnngnigllghshnnlvasswyynnirkntssngcfws 1081

QY 421 FISKHEGMOEN 431  
 DB 1082 fiskhegngen 1092

RESULT 12  
 AAY77138  
 ID AAY77138 standard; Protein; 432 AA.  
 AC AAY77138;  
 DT 08-MAY-2000 (first entry)  
 DE Synthetic botulinum neurotoxin serotype F (BONTF) C-terminal fragment.  
 KW Botulinum neurotoxin; heavy chain; BONT; serotype F;  
 KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
 KW VEE; botulinism; vaccine; diagnosis; drug screening.  
 OS Clostridium botulinum.  
 SX Synthetic.  
 PN WO200002524-A2.  
 PD 20-JAN-2000.  
 PF 09-JUL-1999; 99WO-US15570.  
 PR 10-JUL-1998; 98US-0092416.  
 PR 12-MAY-1999; 99US-0133870.  
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
 DR WPI: 2000-160827/14.  
 DR N-PSDB; AA87216.  
 PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 toxin serotypes A-G, is used for inducing an immune response against  
 botulinum -  
 PS Claim 27; Page -: 54pp; English.

The invention relates to novel vaccines that induce a protective immune response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F and G (BONTA-BONTG). The vaccine of the invention is novel recombinant DNA construct comprising a vector, and at least one nucleic acid fragment comprising a C-terminal heavy chain fragment (Hc) from BONT serotypes A-G. In preferred embodiments of the invention, the vector is a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this vector results in the production of large amounts of a protein encoded by a sequence cloned into the replicon. The constructs are used to produce vaccines against botulinism. The proteins can also be used as diagnostic tools for the diagnosis of botulinism. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulinism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AAY77134-177139 represent synthetic BONT Hc fragments used in the present invention. The DNA encoding these sequences had been optimised for codon usage for expression in yeast. Note: This sequence is not given in the specification, but is decoded from the BONTF Hc DNA sequence given on pages 45-46.

Sequence 432 AA;  
 Query Match 99.3%; Score 2271; DB 21; Length 432;

Best Local Similarity 99.3%; Pred. No. 1,6e-166;  
 Matches 428; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYTNDKILILYFNKLRKIKRDKNSITDREYNNKFDISGSGNSISINGDIYSTRNOF 60  
 DB 2 syndklllyfnklykkkdknsildmryennkfddisgsgnsisngdyystnrgf 61

QY 61 GYSSKPEVNIQNDIYNGRONFSIFWRIIPKYNVNNINNYTTIDCIRNNNSG 120  
 DB 62 gyskpevniqndiyngrnfsifwriipkynvnninnyttidcirmnsg 121

QY 121 WKISLNNKTIWLPDAGNCKLVFNWTOMISISPYTKWTFPTIRNNLGSRTYNG 180  
 DB 122 wkislwnktilwlpdagncqlvfnwtomisisytkwtfptirnnlgsrtyng 181

QY 181 NLDEKSIWLGDIHVSQNLIFKIVGCDNDRYVGIRYKVPDELTETETLYSDPEPP 240  
 DB 182 nldekiswlgdihvsqnlifkivgcdndrvgyirvykvpdelteletysdepdp 241

QY 241 SILKDFWGNLYLNKRRYLLNLRTDKSITQNSNPLINQGRGYOKPNIFSNRLYTCV 300  
 DB 242 silkdfigwnyllynkryllnlrtkdsitqnsnplinqgrgyokpnlfsnrltylcv 301

QY 301 EVIIRKNGSTDISNTDNEFRKNDLAYINVDREYRLYADISIAPKEIKIRTSNSN 360  
 DB 302 eviirknsgstdisntdnefrkndlayinvdreyrlyadisiapkeikirtsnsn 361

QY 361 NSLGOIIVMDSIGNCTNFMQNNNGNIGLGFHSNNLVASSWYNNIRKNTSNCCEFS 420  
 DB 362 nslgoiivmdsignctnmfmqnnngniglgfhsnnlvasswyynnirkntsncfcfs 421

QY 421 FISKHEGMOEN 431  
 DB 422 fiskhegngen 432

RESULT 13  
 ID AAM68399 standard; Protein; 448 AA.  
 AC AAM68399;  
 DT 07-DEC-1998 (first entry)

Clostridium botulinum type F toxin C fragment.  
 KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;  
 KW botulinism; BoTF.  
 OS Clostridium botulinum serotype F strain 202F (ATCC 23387).  
 SX Synthetic.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "N-terminal His tag"

MO9808540-A1.  
 05-MAR-1998.  
 28-AUG-1997; 97WO-US15394.  
 28-AUG-1996; 96US-0704159.  
 (OPH1-) OPHIDIAN PHARM INC.  
 Thallay BS, Williams JA;  
 WPI: 1998-230234/20.  
 DR N-PSDB; AAV30593.  
 Host cell containing recombinant expression vector encoding Clostridium botulinum type B or E toxin - useful to treat humans



Db 266 tnlldfwgnylllydkeyyllnvlpknfidrtkdslnimirs-----tllanrls 320  
 QY 299 GVEVIRK--NGSTDISTDNFVRKNDLAYIN-VVDROVEYRLADISAKPEKTIKLR 355  
 Db 321 gikvkiqrvnssn---dnlvkrndqylnfvaskthlfpdyadatlcnkcklkl-- 374  
 QY 356 TSNSNSLSGQIIVMSDIGNCTNPNFONNNGNIGLIGFHSNNLVASSWYNNIRKNTSSN 415  
 Db 375 -sssgnrlngvymvsvgnctmfnknngnigllgfkadvastwyltmrdhtsn 433  
 QY 416 GCFWSTISKHGMOE 430  
 Db 434 gcfwnflseehgwqe 448  
 RESULT 15  
 AAB04094 standard; Protein; 449 AA.  
 XX AAB04094;  
 DE 11-APR-2001 (first entry)  
 DE Botulinum toxin heavy chain C-terminal sequence (serotype E).  
 DE Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
 KN recombinant vector; antigen; immune response; vaccine; bacterium;  
 KN infection.  
 XX Synthetic.  
 OS Clostridium botulinum.  
 XX WC200067700-A2.  
 PN 16-NOV-2000.  
 PF 12-MAY-2000; 2000MO-US12890.  
 PR 12-MAY-1999; 99US-0133865.  
 PR 12-MAY-1999; 99US-0133866.  
 PR 12-MAY-1999; 99US-0133867.  
 PR 12-MAY-1999; 99US-0133868.  
 PR 12-MAY-1999; 99US-0133869.  
 PR 12-MAY-1999; 99US-0133873.  
 PR 29-JUL-1999; 99US-0146192.  
 PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
 DR WPI: 2001-016048/02.  
 DR N-PSDB; AAM54488.  
 XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 PT vaccine against botulinism  
 XX  
 PS Claim 3; Fig 7b; 73pp; English.  
 CC Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 CC chain and then posttranslationally nicked, forming a dimeric  
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 CC remain linked by a disulfide bond. Nucleic acids encoding the  
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 CC expression vectors and expressed in transformed cells to produce  
 CC peptide antigens useful for eliciting an immune response to give  
 CC protective immunity against botulinum neurotoxin, which causes  
 CC botulinism. The nucleic acids are expressible in a recombinant  
 CC organisms such as Escherichia coli or Pichia pastoris. The use  
 CC of recombinant nucleic acids are advantageous since it eliminates  
 CC the need to culture large quantities of hazardous toxin-producing  
 CC bacterium. Production yield from the genetically engineered product

75

CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.  
 XX  
 SQ Sequence 449 AA:  
 Query Match 63.4%; Score 1451.5; DB 22; Length 449;  
 Best Local Similarity 63.0%; Pred. No. 17e-103;  
 Matches 274; Conservative 74; Mismatches 70; Indels 17; Gaps 7;  
 QY 1 SYTDKLLIYENKLYKRIKIDNSITLDMRYENKFDISGYSNISINGDYIYSTNRQOF 60  
 Db 26 sytdcklllsyfnkffkrikssvlnmrykndyvtsgysnlnlndgdykypnkqf 85  
 QY 61 GYSSKPEEYVNIQANNNDIYNGRYONFSEISFWYRIPKPYFNK-VNLNNEYTIIDCIRNNNS 119  
 Db 86 gylndckltelnlsqndyillydnkykntsfwriipnydnkivvneyellncmrdns 145  
 QY 120 GWRISLNYKRIIWTLODFRAGNNOKLVEPYTOMISIDYINRWIRFVTITNRLGNSRIYIN 179  
 Db 146 gwkslnhneliwtlqdnaglnqklatfngnanglsdylnkwtlvtlndrlgdslyln 205  
 QY 180 GNLIDKSTSNIGDIHVSNDNLEFKYGCNDTRYGIRYKVEDTELKTEIETIVSDPD 239  
 Db 206 gnlldqkslnlgnlvsdnllfkivncsytylglyfnlfdcldeletqlylgnpn 265  
 QY 240 PSILKDPGWYLLYKRYLLNLRTDKSI-TQNSWFLNINQOGYQKPNFSTNRXT 298  
 Db 266 tnlldfwgnylllydkeyyllnvlpknfidrtkdslnimirs-----tllanrls 320  
 QY 299 GVEVIRK--NGSTDISTDNFVRKNDLAYIN-VVDROVEYRLADISAKPEKTIKLR 355  
 Db 321 gikvkiqrvnssn---dnlvkrndqylnfvaskthlfpdyadatlcnkcklkl-- 374  
 QY 356 TSNSNSLSGQIIVMSDIGNCTNPNFONNNGNIGLIGFHSNNLVASSWYNNIRKNTSSN 415  
 Db 375 -sssgnrlngvymvsvgnctmfnknngnigllgfkadvastwyltmrdhtsn 433  
 QY 416 GCFWSTISKHGMOE 430  
 Db 434 gcfwnflseehgwqe 448

Search completed: August 8, 2002, 09:42:45  
 Job time: 219 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:13:13 ; Search time 24.69 Seconds

(without alignments)  
675,906 Million cell updates/sec

Title: US-08-981-087a-1

Perfect score: 431

Sequence: 1 STNDKILILYFNKLYKRIK.....TSSNGCFWFSFKERHGOEN 431

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size: 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	6.3	1274	1	BXF_CLOBO
2	13	3.3	1250	1	BXF_CLOBO
3	15	3.5	1250	1	BXF_CLOBO
4	11	2.6	1235	1	BXA1_CLOBO
5	11	2.6	1235	1	BXA2_CLOBO
6	10	2.3	458	1	MB22_ARATH
7	8	1.9	1230	1	BXA2_CLOBO
8	8	1.9	1236	1	BXA2_CLOBO
9	7	1.6	1236	1	TDX2_THBAC
10	7	1.6	267	1	EUTT_ECOLI
11	7	1.6	267	1	EUTT_ECOLI
12	7	1.6	358	1	EUTT_ECOLI
13	7	1.6	358	1	VAL1_CLYN
14	7	1.6	387	1	VAL1_CLYN
15	7	1.6	387	1	FIR2_ADE41
16	7	1.6	432	1	FIR2_ADE41
17	7	1.6	432	1	YZZ1_METJA
18	7	1.6	441	1	MGTA_THEMA
19	7	1.6	449	1	MURF_RICPR
20	7	1.6	501	1	VGIC_HSYMB
21	7	1.6	501	1	VGIC_HSYMB
22	7	1.6	505	1	VGIC_HSYMB
23	7	1.6	508	1	YMO5_HSYMB
24	7	1.6	511	1	YMO5_HSYMB
25	7	1.6	515	1	CRPA_DROME
26	7	1.6	523	1	CRPA_DROME
27	7	1.6	608	1	CRPA_DROME
28	7	1.6	608	1	EDD_HELPJ
29	7	1.6	608	1	EDD_HELPJ
30	7	1.6	608	1	EDD_HELPJ
31	7	1.6	614	1	PARC_MYCE
32	7	1.6	614	1	PARC_MYCE
33	7	1.6	944	1	AMPN_MANS
					Y166_UREPA

34	7	1.6	1035	1	TTA9_HUMAN	013797 homo sapien
35	7	1.6	1132	1	PHY1_PUYP	P36505 physcomitre
36	7	1.6	1132	1	PHY1_PUYP	049413 mycoplasma
37	7	1.6	1276	1	BXD_CLOBO	P19321 clostridium
38	7	1.6	1655	1	OMP_RICCN	094633 r outer mem
39	7	1.6	1656	1	OMP_RICCN	006633 r outer mem
40	7	1.6	2292	1	FOLD_EMCYD	P17593 encyphatony
41	7	1.6	2292	1	FOLD_EMCYD	P17594 encyphatony
42	6	1.4	48	1	ATP8_ASPPA	P00857 aspergillus
43	6	1.4	48	1	ATP8_ASPPA	P00857 aspergillus
44	6	1.4	49	1	Y195_BPT7	P03804 bacterioph
45	6	1.4	63	1	RL25_BUCAR	P46174 buchnera ap

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	1274 AA.
1	BXF_CLOBO			
AC	P30996;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Boculinum neurotoxin type F precursor (EC 3.4.24.69) (BONT/F)			
DE	(Bontolixysin F).			
GN	BONT.			
OS	Clostridium botulinum.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1491;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=ATCC 23387;			
RX	MEDLINE=93012902; PubMed=1398040;			
RA	East A.K., Richardson P.T., Allway D., Collins M.D.,			
RA	Roberts T.A., Thompson D.E.,			
RT	Sequence of the gene encoding type F neurotoxin of Clostridium			
RT	botulinum.			
RL	FEWS Microbiol. Lett. 75:225-230(1992).			
RN	[2]			
RP	SEQUENCE OF 1-64 FROM N.A.			
RP	STRAIN=HOBBS FT10;			
RX	MEDLINE=94297488; PubMed=7764998;			
RA	East A.K., Collins M.D.;			
RA	"Conserved structure of genes encoding components of botulinum			
RT	neurotoxin complex M and the sequence of the gene coding for the			
RT	neurotoxin component in nonproteolytic Clostridium botulinum type F.";			
RL	Curr. Microbiol. 29:69-77(1994).			
RN	[3]			
RP	SEQUENCE OF 634-1002 FROM N.A.			
RP	MEDLINE=94013372; PubMed=8408542;			
RX	Campbell K., East A.K., Collins M.D.;			
RA	"Gene probes for identification of the botulinum neurotoxin gene and			
RT	specific identification of neurotoxin types B, E, and F.";			
RL	J Clin. Microbiol. 31:2255-2262(1993).			
RN	[4]			
RP	IDENTIFICATION OF SUBSTRATE.			
RP	MEDLINE=94230352; PubMed=8175689;			
RX	Yanase B., Ryke E.M., Suedhof T.C., Jahn R., Niemann H.,			
RA	Rouges B., Ryke E.M., Suedhof T.C., Jahn R., Niemann H.,			
RT	"Cleavage of members of the synaptobrevin/VAMP family by types D and			
RT	F botulinum neurotoxins and tetanus toxin.";			
RL	J Biol. Chem. 269:12764-12772(1994).			
RN	[5]			
RP	FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER			
RP	RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED			
RP	AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD			
RP	WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT			
RP	INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC			
RP	ENDOPETIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-1-LYS-59			
RP	BOND OF SYNAPTOSOMALIN-1 AND -2.			
RP	FUNCTION: CATALYTIC ACTIVITY: limited hydrolysis of proteins of the			
RP	neurotoxicity apparatus, synaptobrevins, SNAP25 or syntaxin. No			

CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M92906; AAA23263.1; -;  
 CC EMBL; S73676; AAC60475.1; -;  
 CC EMBL; X70820; CAA50151.1; -;  
 CC EMBL; X70816; CAA50147.1; -;  
 CC HSSP; P10845; 3BTA.  
 CC MEROPS; M27.002; -;  
 CC InterPro; IPR000395; Bontoxilysin.  
 CC InterPro; IPR000130; Zn\_Mtpeptidse.  
 CC Pfam; PF01742; Peptidase\_M27; 1.  
 CC PRINTS; PR00760; BONTOTOXILYSIN.  
 CC ProDom; PD001963; Bontoxilysin; 1.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 CC CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.  
 CC CHAIN 437 1274 BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.  
 CC METAL 227 227 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC AC SITE 228 228 BY SIMILARITY.  
 CC METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC DISULF 429 445 INTERCHAIN (PROBABLE).  
 CC SEQUENCE 1274 AA; 146709 MW; 5B99756A7438B921 CRC64;  
 CC  
 CC Query Match 6.3%; Score 27; DB 1; Length 1274;  
 CC Best Local Similarity 100.0%; Pred. No. 2,1e-19;  
 CC Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 157 YINKWIVTITNNRLGNSRIYINGMLI 183  
 CC ||||||||||||||||||||||||||||  
 CC Db 1006 YINKWIVTITNNRLGNSRIYINGMLI 1032  
 CC  
 CC RESULT 2  
 CC EXE\_CLOBO STANDARD; PRT; 1250 AA.  
 CC ID EXE\_CLOBO  
 CC AC Q00496;  
 CC DT 01-JUL-1993 (Rel. 26, Created)  
 CC DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 CC DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 CC DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)  
 CC (Bontoxilysin E).  
 CC OS Clostridium botulinum.  
 CC OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 CC Clostridium.  
 CC NCBI\_TaxID=1491;  
 CC RX [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=BEUGA;  
 CC RA MEDLINE=92181428; PubMed=1543481;  
 CC "Sequences of the botulinum neurotoxin E derived from Clostridium  
 CC botulinum type E (Strain Beuga) and Clostridium butyricum (strains  
 CC ATCC 43181 and ATCC 43755).";  
 CC RL Biochem. Biophys. Res. Commun. 183:107-113(1992).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.

RX MEDLINE=92174922; PubMed=1541280;  
 RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;  
 RT "The complete amino acid sequence of the Clostridium botulinum type-E  
 RT neurotoxin, derived by nucleotide-sequence analysis of the encoding  
 RT gene."; Biochem. 204:657-667(1992).  
 RL Eur. J. Biochem. 204:657-667(1992).  
 RN [3]  
 RP SEQUENCE OF 1-251 FROM N.A.  
 RX MEDLINE=90264400; PubMed=2160960;  
 RA Bliz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;  
 RT "The complete sequence of botulinum neurotoxin type A and comparison  
 RT with other clostridial neurotoxins."; J. Biol. Chem. 265:9153-9158(1990).  
 RL [4]  
 RN SEQUENCE OF 1-13.  
 RX MEDLINE=85197963; PubMed=3888113;  
 RA Schmidt J.J., Sathymoorthy V., Dasgupta B.R.;  
 RT "Partial amino acid sequences of botulinum neurotoxins types B and  
 RT E."; Arch. Biochem. Biophys. 238:544-548(1985).  
 RL Arch. Biochem. Biophys. 238:544-548(1985).  
 RN [5]  
 RP SEQUENCE OF 419-426.  
 RX MEDLINE=90344918; PubMed=2116911;  
 RA Gimenez J.A., Dasgupta B.R.;  
 RT "Botulinum neurotoxin type E, fragmented with endoproteinase Lys-C  
 RT reveals the site trypsin nicks and homology with tetanus  
 RT neurotoxin."; Biochimie 72:213-217(1990).  
 RL [6]  
 RN IDENTIFICATION OF SUBSTRATE.  
 RP MEDLINE=94063091; PubMed=8243676;  
 RX Schlavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
 RA Benfenati F., Wilson M.C., Montecucco C.;  
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 RT COOH-terminal peptide bonds."; FEBS Lett. 335:99-103(1993).  
 RL [7]  
 RN IDENTIFICATION OF SUBSTRATE.  
 RP MEDLINE=94124495; PubMed=8294407;  
 RX Blinz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 RA Jahn R., Niemann H.;  
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins."; J. Biol. Chem. 269:1617-1620(1994).  
 RL [8]  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-  
 CC 181 BOND IN SNAP-25.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X62089; CAA43999.1; -;  
 CC EMBL; X62683; CAA44558.1; -;  
 CC PIR; A60027; A60027.

```

DR PIR: B35294; B35294.
DR PIR: JH0257; JH0257.
DR PIR: S08575; S08575.
DR PIR: S1811; S1811.
DR PIR: S21178; S21178.
DR HSP: P10845; 3BTA.
DR MEROPS: M27.002; -.
DR InterPro: IPR000395; Bontoloxilysin.
DR InterPro: IPR000130; Zn_MTEPtease.
DR Pfam: PF01742; Peptidase_M27.1.
DR PRINTS: PR00760; BONTOLIXYSIN.
DR PRODOM: PD001963; Bontoloxilysin.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
KM INIT_MET 0
FT CHAIN 1 421
FT CHAIN 422 1250
FT METAL 211 211
FT ACT_SITE 212 212
FT METAL 215 215
FT DISULFID 411 425
FT CONFLICT 197 176
FT CONFLICT 197 197
FT CONFLICT 339 339
FT CONFLICT 772 772
FT CONFLICT 962 966
FT CONFLICT 1194 1194
SQ SEQUENCE 1250 AA; 143712 MW; D9FCE26DDA041EB4 CRC64;

Query Match 3.5%; Score 15; DB 1; Length 1250;
Best Local Similarity 100.0%; Pred. No. 4.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ISDYINKMIFVTITN 168
DB 982 ISDYINKMIFVTITN 996

RESULT 3
BXAL_CLOBO STANDARD; PRT: 1250 AA.
ID BXAL_CLOBO
AC P30995;
DR 01-JUL-1993 (Rel. 26, Created)
DR 01-JUL-1993 (Rel. 26, Last sequence update)
DR 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)
DE (Bontoloxilysin E).
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43181, AND ATCC 43755;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE OF 1-251 FROM N.A.
RC STRAIN-BL6340;
RX MEDLINE=91237316; PubMed=2033376;
RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA Yokosawa N., Yashiki T., Oguma K.;
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. Gen. Microbiol. 137:519-525(1991).
RN [3]
RP SEQUENCE OF 1-48.

```

```

RC STRAIN=5262;
RA Gimenez J., Foley J., Dasgupta B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT partial sequence and comparison.";
RL FASEB J. 2:41750-41750(1988).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -1- CATALYTIC ACTIVITY: limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62088; CAA43998.1; -.
DR EMBL: X53180; CAA37321.1; -.
DR PIR: JH0256; JH0256.
DR PIR: S16145; S16145.
DR HSP: P10845; 3BTA.
DR MEROPS: M27.002; -.
DR InterPro: IPR00395; Bontoloxilysin.
DR InterPro: IPR00130; Zn_MTEPtease.
DR Pfam: PF01742; Peptidase_M27.1.
DR PRINTS: PR00760; BONTOLIXYSIN.
DR PRODOM: PD001963; Bontoloxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
KM INIT_MET 0
FT CHAIN 1 421
FT CHAIN 422 1250
FT METAL 211 211
FT METAL 212 212
FT METAL 215 215
FT DISULFID 411 425
FT CONFLICT 229 229
SQ SEQUENCE 1250 AA; 143265 MW; 8171B5B2C312857 CRC64;

Query Match 3.5%; Score 15; DB 1; Length 1250;
Best Local Similarity 100.0%; Pred. No. 4.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ISDYINKMIFVTITN 168
DB 982 ISDYINKMIFVTITN 996

RESULT 4
BXAL_CLOBO STANDARD; PRT: 1295 AA.
ID BXAL_CLOBO
AC P10845; P10839; F01561.
DR 01-JUL-1989 (Rel. 11, Created)
DR 01-JUL-1993 (Rel. 26, Last sequence update)
DR 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
DE (Bontoloxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, Light-

```

DE chain: Botulinum neurotoxin A, heavy-chain).  
 GN BOTA OR BNA OR ATX.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIC 2916;  
 RX MEDLINE=90235864; PubMed=2185020;  
 RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,  
 RA Shone C.C., Atkinson T., Melling J., Minton N.P.;  
 RT "The complete amino acid sequence of the Clostridium botulinum type A  
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding  
 RT gene.";  
 RL Eur. J. Biochem. 189:73-81(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=62A;  
 RX MEDLINE=90264400; PubMed=2160960;  
 RA Binz B., Kuatzono H., Wille M., Frevent J., Wernars K., Niemann H.;  
 RT "The complete sequence of botulinum neurotoxin type A and comparison  
 RT with other clostridial neurotoxins.";  
 RL J. Biol. Chem. 265:9153-9158(1990).  
 RN [3]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RC STRAIN=62A;  
 RX MEDLINE=97016817; PubMed=8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RT "Organization and phylogenetic interrelationships of genes encoding  
 RT components of the botulinum toxin complex in proteolytic Clostridium  
 RT botulinum types A, B, and F: evidence of chimeric sequences in the  
 RT gene encoding the nontoxic nonhemagglutinin component.";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 RN [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RC STRAIN=HALL;  
 RX MEDLINE=89350959; PubMed=2669749;  
 RA Betley M.J., Somers E., Dasgupta B.R.;  
 RT "Characterization of botulinum type A neurotoxin gene: delineation of  
 RT the N-terminal encoding region.";  
 RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).  
 RN [5]  
 RP SEQUENCE OF 1-18 FROM N.A.  
 RC STRAIN=TYPE A NIH;  
 RX MEDLINE=96096783; PubMed=8521962;  
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
 RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
 RT components of Clostridium botulinum type A progenitor toxins.";  
 RL FEBS Lett. 376:41-44(1995).  
 RN [6]  
 RP SEQUENCE OF 1-16.  
 RX MEDLINE=84178501; PubMed=6370252;  
 RA Schmidt J.J., Sartyoorthy V., Dasgupta B.R.;  
 RT "Partial amino acid sequence of the heavy and light chains of  
 RT botulinum neurotoxin type A.";  
 RL Biochem. Biophys. Res. Commun. 119:900-904(1984).  
 RN [7]  
 RP SEQUENCE OF 1-46.  
 RA Dasgupta B.R., Foley J., Niece R.;  
 RT "Partial sequence of the light chain of botulinum neurotoxin type A.";  
 RL Biochemistry 26:4162-4162(1987).  
 RN [8]  
 RP SEQUENCE OF 1-5 AND 444-456.  
 RX MEDLINE=91120847; PubMed=2126206;  
 RA Dasgupta B.R., Dekleva M.L.;  
 RT "Botulinum neurotoxin type A: sequence of amino acids at the  
 RT N-terminus and around the nicking site.";  
 RL Biochimie 72:661-664(1990).  
 RN [9]  
 RP SEQUENCE OF 448-464 AND 872-895.  
 RX MEDLINE=89024662; PubMed=3178218;  
 RA Sathyamoorthy V., Dasgupta B.R., Foley J., Niece R.L.;

RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two  
 RT halves and their partial sequences.";  
 RL Arch. Biochem. Biophys. 266:142-151(1988).  
 RN [10]  
 RP SEQUENCE OF 448-482.  
 RX MEDLINE=85285016; PubMed=3896784;  
 RA Shone C.C., Hambleton P., Melling J.;  
 RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin  
 RT and purification of two tryptic fragments. Proteolytic action near  
 RT the COOH-terminus of the heavy subunit destroys toxin-binding  
 RT activity.";  
 RL Eur. J. Biochem. 151:75-82(1985).  
 RN [11]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94063091; PubMed=8243676;  
 RA Schiavo G., Santucci A., Dasgupta B.R., Melia P.P., Jontes J.,  
 RA Benfenati F., Wilson M.C., Montecucco C.;  
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 RT COOH-terminal peptide bonds.";  
 RL FEBS Lett. 335:99-103(1993).  
 RN [12]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=944124495; PubMed=8294407;  
 RA Binz T., Blas J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 RA Jahn R., Niemann H.;  
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";  
 RL J. Biol. Chem. 269:1617-1620(1994).  
 RN [13]  
 RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.  
 RX PubMed=11700044;  
 RA Biondi M., Cacchi P., Johnson E.A., Montecucco C., Rossetto O.;  
 RT "Site-directed mutagenesis identifies active-site residues of the  
 RT light chain of botulinum neurotoxin type A.";  
 RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).  
 RN [14]  
 RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).  
 RX MEDLINE=98455071; PubMed=9783750;  
 RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;  
 RT "Crystal structure of botulinum neurotoxin type A and implications  
 RT for toxicity.";  
 RL Nat. Struct. Biol. 5:898-902(1998).  
 CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin  
 CC binds with high affinity to peripheral neuronal presynaptic  
 CC membrane, is then internalized by receptor-mediated endocytosis.  
 CC The C-terminus of the heavy chain (H) is responsible for the  
 CC adherence of the toxin to the cell surface while the N-terminus  
 CC mediates transport of the light chain from the endocytic vesicle  
 CC to the cytosol. After translocation, the light chain (L)  
 CC hydrolyzes the 197-Gln-1-Arg-198 bond in SNAP-25, thereby blocking  
 CC neurotransmitter release. Inhibition of acetylcholine release  
 CC results in flaccid paralysis, with frequent heart or respiratory  
 CC failure.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a  
 CC heavy chain (H).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- PHARMACEUTICAL: Available under the name BOTOX(R) (Allergan) for  
 CC the treatment of strabismus and blepharospasm associated with  
 CC dystonia and cervical dystonia. Also used for the treatment of  
 CC hemifacial spasm and a number of other neurological disorders  
 CC characterized by abnormal muscle contraction.  
 CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of  
 CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@slsb.ch)
CC EMBL: X52066; CAA36289.1; -
DR EMBL: M30196; AAA23262.1; -
DR EMBL: X92973; CAA63551.1; -
DR EMBL: D67030; BAA1051.1; -
DR EMBL: M27892; AAA23269.1; -
DR PIR: A35294; B7CLAB.
DR PIR: S09492; S09492.
DR PDB: 3BTA; 01-OCT-99.
DR MEROPS: M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOXILYSIN.
DR PRODOM: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW Pharmacological: 3D-structure.
FT INIT_MET 0
FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT METAL 222 222 ZINC (CATALYTIC).
FT ACT_SITE 223 223 ZINC (CATALYTIC).
FT METAL 226 226 ZINC (CATALYTIC).
FT METAL 261 261 ZINC (CATALYTIC).
FT DISULFID 429 453 INTERCHAIN.
FT DISULFID 1234 1279
FT TRANSMEM 626 646
FT TRANSMEM 655 675
FT TRANSMEM 26 26
FT METAL 261 261
FT METAL 265 265
FT METAL 365 365
FT CONFLICT 1 1
FT CONFLICT 479 479
FT CONFLICT 875 875
FT CONFLICT 891 891
FT CONFLICT 891 891
SO SEQUENCE 1295 AA; 149322 MW; 858342P54862579 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 1295;
Best Local Similarity 100.0%; Pred. No. 0.0065;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 WIFVITNNRL 171
Db 1013 WIFVITNNRL 1023

RESULT 5
BXA2_CLOBO STANDARD; PRT: 1295 AA.
ID BXA2_CLOBO
AC 045894; P77780;
DT 01-MAR-2002 (rel. 41, Created)
DT 01-MAR-2002 (rel. 41, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (BC 3.4.24.69) (BONT/A)
DE (Bontoxilysin A) (BOTOX) [contains: Botulinum neurotoxin A, light-
chain; Botulinum neurotoxin A, heavy-chain].
GN BOTA OR BNA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_Taxid=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Kyoto-F;
RC MEDLINE=94143603; PubMed=8310180;
RA Williams A., East A.K., Lawson P.A., Collins M.D.;
RT "Sequence of the gene coding for the neurotoxin of Clostridium
botulinum type A associated with infant botulism: comparison with

```

```

RT other clostridial neurotoxins";
RL Res. Microbiol. 144:547-556(1993).
RN (2)
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN-Kyoto-F;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC binds with high affinity to peripheral neuronal presynaptic
CC membrane, is then internalized by receptor-mediated endocytosis.
CC The C-terminus of the heavy chain (H) is responsible for the
CC adherence of the toxin to the cell surface while the N-terminus
CC mediates transport of the light chain from the endocytic vesicle
CC to the cytosol. After translocation, the light chain (L)
CC hydrolyzes the 197-Gln-1-Arg-198 bond in SNAP-25, thereby blocking
CC neurotransmitter release. Inhibition of acetylcholine release
CC results in flaccid paralysis, with frequent heart or respiratory
CC failure (by similarity).
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevin, SNAP25 or syntaxin. NO
CC detected action on small molecule substrates.
CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H) (by similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: types A, B, C1, D, E, F, and G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@slsb.ch).
CC
DR EMBL: X73423; CAA51824.1; -.
DR EMBL: X87974; CAA61234.1; -.
DR HSSP: P10845; 3BTA.
DR MEROPS: M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOXILYSIN.
DR PRODOM: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0
FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 223 223 BY SIMILARITY.
FT METAL 226 226 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 261 261 INTERCHAIN (BY SIMILARITY).
FT DISULFID 429 453 BY SIMILARITY.
FT DISULFID 1234 1279
FT TRANSMEM 626 646
FT TRANSMEM 655 675
FT SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;

Query Match 2.6%; Score 11; DB 1; Length 1295;
Best Local Similarity 100.0%; Pred. No. 0.0065;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 WIFVITNNRL 171
Db 1013 WIFVITNNRL 1023

```

```

RESULT 6
MB22_ARATH STANDARD; PRT; 458 AA.
ID MB22_ARATH
AC 080950;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myrosinase binding protein-like At2g39310.
GN At2g39310 OR T16B24.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Bueli C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
CC -1- SIMILARITY: BELONGS TO THE JACALIN LECTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AC004697; AAC28979.1; -
CC DR HSSP; P18674; 1JOT.
CC DR InterPro: IPR001229; Jacalin.
CC DR Pfam: PF01419; Jacalin; 3.
CC DR Lectin; Repeat; Multigene family.
CC KW SEQUENCE 458 AA; 50463 MW; EB01A410563EAA8 CRC64;
CC SQ

```

```

Query Match 2.3%; Score 10; DB 1; Length 458;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 285 YOKPNIFSNT 294
DB 238 YOKPNIFSNT 247

```

---

```

RESULT 7
BXB_CLOBO STANDARD; PRT; 1290 AA.
ID BXB_CLOBO
AC P10844; P10843;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BONT/B)
DE (Botolixylsin B).
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92384550; PubMed=1514783;

```

---

```

RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA Minton N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence."
RT Appl. Environ. Microbiol. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCIC 7273;
RA Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL Submitted (APR-1992) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCIC 7273;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F."
RT J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE=89000987; PubMed=3139097;
RA Dasgupta B.R., Datta A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin."
RL Biochimie 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE=85197963; PubMed=388113;
RA Schmidt J.J., Sathyanoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E."
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [6]
RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE=93054694; PubMed=1429690;
RA Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT "Botulinum neurotoxins are zinc proteases."
RL J. Biol. Chem. 267:23479-23483(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=93063293; PubMed=1331807;
RA Schiavo G., Benfenati F., Poullain B., Rossetto O., de laureto P.P.,
RA Dasgupta B.R., Montecucco C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin."
RL Nature 359:832-835(1992).
RN [8]
RP FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-1-PHE-77 BOND OF
CC SYNAPTOBREVIN-2.
CC [9]
CC CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC [10]
CC SUBUNIT: DISULFIDE-LINKED HETEROOLIGOMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC [11]
CC SUBCELLULAR LOCATION: Secreted.
CC [12]
CC MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC [13]
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entites requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 DR EMBL; M6186; AAA2221.1; -;  
 DR EMBL; Z1934; CAAT799.1; -;  
 DR EMBL; X70817; CAA50148.1; -;  
 DR PIR; S07128; S07128;  
 DR PIR; S07155; S07155;  
 DR PIR; S08562; S08562;  
 DR PIR; S08573; S08573;  
 DR PIR; S08574; S08574;  
 DR PIR; A48940; A48940;  
 DR HSSP; P10845; 3BTA;  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR000130; Zn\_MTPeptide.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 KW INIT\_MET 0  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.  
 FT METAL 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.  
 FT ACT\_SITE 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 436 445 INTERCHAIN (PROBABLE).  
 FT CONFLICT 29 25 T->M (IN REF. 4).  
 FT CONFLICT 217 217 R->G (IN REF. 2).  
 FT CONFLICT 224 224 A->S (IN REF. 2).  
 FT CONFLICT 463 463 S->R (IN REF. 4).  
 SQ SEQUENCE 1290 AA; 150670 MW; D21746ZC024DF43 CRC64;  
 QY 117 NNSGKIS 124  
 Db 957 NNSGKIS 964  
 RESULT 8  
 ID BXG\_CLOBO STANDARD; PRT; 1296 AA.  
 AC 060393;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BONT/G)  
 GN BONG.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=113 / 30;  
 RC MEDLINE=94092745; PubMed=8268233;  
 RA Campbell K., Collins M.D., East A.K.;  
 RT Nucleotide sequence of the gene coding for Clostridium botulinum  
 RT (Clostridium argentineense) type G neurotoxin: genealogical comparison  
 RT with other clostridial neurotoxins.";  
 RL Blochm. Biophys. Acta 1216:487-491(1993).  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE.

CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted (by similarity).  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entites requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC -----  
 DR EMBL; X74162; CAA52275.1; -;  
 DR HSSP; P10845; 3BTA.  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR000130; Zn\_MTPeptide.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 KW INIT\_MET 0  
 FT CHAIN 1 441 BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.  
 FT METAL 442 1296 BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.  
 FT ACT\_SITE 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 435 449 INTERCHAIN (PROBABLE).  
 SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F65C31 CRC64;  
 QY 154 ISDYINKW 161  
 Db 1001 ISDYINKW 1008  
 RESULT 9  
 ID TDXX2\_THERAC STANDARD; PRT; 199 AA.  
 AC 09HJL3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable peroxyliredoxin 2.  
 GN TA0954.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;  
 OC Thermoplasmata.  
 OX NCBI\_TaxID=2303;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RC MEDLINE=20479972; PubMed=11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mews H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermocacidophilic scavenger Thermoplasma  
 RT acidophilum.";  
 RT Nature 407:508-513(2000).  
 CC -1- SIMILARITY: BELONGS TO THE AHPG/TSA FAMILY. TDXX SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AL445066; CAC12083.1; -  
DR InterPro; IPR000866; AbpC-TSA.  
DR Pfam; PF00578; AbpC-TSA; 1.  
KW Antioxidant; Complete proteome.  
FT ACT SITE 40  
FT SITE 40 BY SIMILARITY  
SQ SEQUENCE 199 AA; 22581 MW; 4E266162F5D58162 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 199;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 NIDEXS 187  
|111111|  
Db 104 NIDEXS 110

RESULT 10  
EUTL ECOLI STANDARD; PRT; 267 AA.  
AC P/6554;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).  
GN EUTL OR B2459.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
CC -1- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + cob(II)alamin + H(2)O = phosphate +  
CC diphosphate + adenosylcobalamin.  
CC -1- PATHWAY: ETHANOLAMINE UTILIZATION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AE000332; AAC75512.1; -  
DR Ecogene; EG14189; eutT.  
KW Transferase; Complete proteome.  
SQ SEQUENCE 267 AA; 30171 MW; B51EDAB528B4FA76 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 267;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 LNLNLT 265  
|111111|

Db 199 LNLNLT 205

RESULT 11  
EUTL SALTY STANDARD; PRT; 267 AA.  
AC Q9ZEV4;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).  
GN EUTL OR STM2467.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmomella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L72;  
RX MEDLINE=99395039; PubMed=10464203;  
RA Kofoed E.C., Rappleye C.A., Stojiljkovic I., Roth J.R.;  
RT "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium  
RT encodes five homologues of carboxysome shell proteins.";  
RL J. Bacteriol. 181:5317-5329(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L72 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Speleth J., Clifton S.W., Latreille P.,  
RA Courtney L., Portwolik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,  
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
CC -1- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + cob(II)alamin + H(2)O = phosphate +  
CC diphosphate + adenosylcobalamin.  
CC -1- PATHWAY: ETHANOLAMINE UTILIZATION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AF093749; AAC78114.1; -  
DR EMBL; AE008811; AAL21361.1; -  
DR StyGene; SG10636; eutT.  
KW Transferase; Complete proteome.  
SQ SEQUENCE 267 AA; 30238 MW; 9502A28FDB4DC9E4 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 267;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 LNLNLT 265  
|111111|  
Db 199 LNLNLT 205

RESULT 12  
VALI\_CLVK STANDARD; PRT; 358 AA.  
AC P14982;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE ALI protein (40.4 kDa protein).



```

GN ACI
OS Cassava latent virus (strain West Kenyan 844).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10818;
RN [1]
RP SEQUENCE FROM N.A.
RA Stanley J., Gay M.R.;
RT Nucleotide sequence of cassava latent virus DNA.
RL Nature 301:260-262(1983).
CC -1- SIMILARITY: BELONGS TO GEMINIVIRUSES ALL PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, J02057; -; NOT_ANNOTATED_CDS.
CC InterPro: IPR001191; Gemini_AL1.
CC Pfam: PF00799; Gemini_AL1.
CC PRINTS: PR00227; GEMCOATL1.
DR ProDom: PD000736; Gemini_AL1.
KM ATP-binding. 220 ATP (POTENTIAL).
FT NP_BIND ED173E753E92D69 CRC64;
SQ SEQUENCE 358 AA; 40346 MW; ED173E753E92D69 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 TITNRL 171
Db 68 TITNRL 74

RESULT 13
ID VAL1_CLVN STANDARD; PRT; 358 AA.
AC P14972;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE AL1 protein (40.4 kDa protein).
GN AC1.
OS Cassava latent virus (strain Nigerian).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10819;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-90174930; PubMed-2308831;
RA Morris B., Coates L., Lowe S., Richardson K., Eddy P.;
RT Nucleotide sequence of the infectious cloned DNA components of
RT African cassava mosaic virus (Nigerian strain).
RL Nucleic Acids Res. 18:197-198(1990).
CC -1- SIMILARITY: BELONGS TO GEMINIVIRUSES ALL PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X17095; CAA34953.1; -.
CC PIR: S07594; S07594.
DR InterPro: IPR001191; Gemini_AL1.
DR Pfam: PF00799; Gemini_AL1.
DR PRINTS: PR00227; GEMCOATL1.
DR ProDom: PD000736; Gemini_AL1.

```

```

KM ATP-binding. 220 ATP (POTENTIAL).
FT NP_BIND 40346 MW; 1DB16BB0CB2D5E2C CRC64;
SQ SEQUENCE 358 AA; 40346 MW; 1DB16BB0CB2D5E2C CRC64;

Query Match 1.6%; Score 7; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 TITNRL 171
Db 68 TITNRL 74

RESULT 14
ID FIB2_ADE40 STANDARD; PRT; 387 AA.
AC P18048;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Fiber protein 2.
OS Human adenovirus type 40.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28284;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DUGAN;
RX MEDLINE-94087748; PubMed-8263936;
RA Davison A.V., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT The DNA sequence of adenovirus type 40.
RL J. Mol. Biol. 234:1308-1316(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-93297140; PubMed-8517033;
RA Kidd A.H., Chroboczek J., Cusack S., Ruigrok R.W.H.;
RT "Adenovirus type 40 virions contain two distinct fibers."
RL Virology 192:73-84(1993).
RN [3]
RP SEQUENCE OF 167-387 FROM N.A.
RX MEDLINE-89370295; PubMed-2773314;
RA Kidd A.H., Erasmus M.J.;
RT "Sequence characterization of the adenovirus 40 fiber gene."
RL Virology 172:134-144(1989).
CC -1- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L19443; AAC13979.1; -.
CC PIR: A40048; ERADY4.
DR InterPro: IPR000939; Adeno_fiber2.
DR InterPro: IPR000978; Adeno_fiber_knob.
DR InterPro: IPR000931; Adeno_fibre.
DR Pfam: PF00541; adeno_fiber2.1.
DR Pfam: PF00608; adeno_fiber2.5.
DR PRINTS: PR00307; ADENOVSFIBRE.
KM Fiber protein.
FT CONFLICT 226 G -> S (IN REF. 2 AND 3).
SQ SEQUENCE 387 AA; 41346 MW; 11A3C1FC061A3ACB CRC64;

Query Match 1.6%; Score 7; DB 1; Length 387;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 358 NSNNSLG 364  
 |||||  
 Db 89 NSNNSLG 95

## RESULT 15

FIB2\_ADEA1 STANDARD; PRT; 387 AA.  
 AC P16883;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Fiber protein 2.  
 OS Human adenovirus type 41.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=10524;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TAK;  
 RX MEDLINE=90245595; PubMed-2336370;  
 RA Pieniazek N.J., Slemenda S.B., Pieniazek D., Velaarde J. Jr.,  
 RA Luftig R.B.;  
 RT "Human enteric adenovirus type 41 (Tak) contains a second fiber  
 RT protein gene.";  
 RL Nucleic Acids Res. 18:1901-1901(1990).  
 RN [2]  
 RP SEQUENCE OF 337-387 FROM N.A.  
 RC STRAIN-FB585;  
 RX MEDLINE=91021015; PubMed-2219717;  
 RA Kidd A.H., Erasmus M.J., Tienessen C.T.;  
 RT "Fiber sequence heterogeneity in subgroup F adenoviruses.";  
 RL Virology 179:139-150(1990).  
 CC -1- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND  
 CC BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.  
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X17016; CA34882.1; -;  
 DR EMBL: M60327; AAA42505.1; -;  
 DR PIR: S09217; ERADN1.  
 DR PIR: A45352; A45352.  
 DR HSSP: P11818; 1KNB.  
 DR InterPro: IPR000939; Adeno\_fiber2.  
 DR InterPro: IPR000978; Adeno\_fiber\_knob.  
 DR InterPro: IPR000931; Adeno\_fibre.  
 DR Pfam: PF00541; adeno\_fiber; 1.  
 DR Pfam: PF00608; adeno\_fiber2; 5.  
 DR PRINTS: PR00307; ADENOVSFIBRE.  
 KW Fiber protein.  
 SQ SEQUENCE 387 AA; 41397 MW; 8652E785276573C7 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 387;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 358 NSNNSLG 364  
 |||||  
 Db 89 NSNNSLG 95

Search completed: August 15, 2002, 11:24:37  
 Job time: 684 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:25 ; Search time 96.53 Seconds  
(without alignments) 165.696 Million cell updates/sec

Title: US-08-981-087A-2

Perfect score: 144  
Sequence: 1 STYNDKILILYFNKILYKKIK.....LNWKIITLLODTAGNCKL 144

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size: 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A\_Geneseq\_032802.\*

```
1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	100.0	144	18	AAW09015
2	144	100.0	431	18	AAW09014
3	144	100.0	432	22	AAW04096
4	144	100.0	432	22	AAW04103
5	144	100.0	645	22	AAW07894
6	144	100.0	660	22	AAW07898
7	144	100.0	685	22	AAW07893
8	144	100.0	862	22	AAW07890
9	144	100.0	887	22	AAW07892
10	144	100.0	1032	22	AAW07901
11	144	100.0	1059	21	AAW33309

12	144	100.0	1084	21	AAW93312
13	144	100.0	1092	22	AAW07900
14	88	61.1	432	21	AAW77138
15	26	18.1	448	19	AAW68399
16	11	7.6	419	22	AAW04095
17	11	7.6	449	21	AAW77137
18	11	7.6	449	22	AAW04094
19	11	7.6	451	19	AAW68396
20	11	7.6	452	19	AAW68395
21	8	5.6	233	21	AAW77143
22	8	5.6	382	21	AAW36303
23	8	5.6	415	22	AAW04083
24	8	5.6	432	21	AAW77142
25	8	5.6	434	22	AAW04090
26	8	5.6	435	22	AAW04088
27	8	5.6	437	22	AAW04088
28	8	5.6	438	17	AAW95008
29	8	5.6	438	19	AAW68389
30	8	5.6	439	21	AAW77134
31	8	5.6	439	22	AAW04085
32	8	5.6	440	21	AAW77135
33	8	5.6	440	22	AAW04091
34	8	5.6	445	19	AAW68391
35	8	5.6	445	17	AAW95009
36	8	5.6	462	17	AAW68390
37	8	5.6	472	19	AAW68394
38	8	5.6	472	19	AAW68393
39	8	5.6	837	21	AAW77140
40	8	5.6	847	22	AAW04081
41	8	5.6	848	22	AAW04082
42	8	5.6	1067	21	AAW93307
43	8	5.6	1070	21	AAW93308
44	8	5.6	1092	21	AAW93310
45	8	5.6	1095	21	AAW93311

#### ALIGNMENTS

```
RESULT 1
AAW09015
ID: AAW09015 standard: Protein: 144 AA.
XX
XX AAW09015:
XX
XX 31-MAR-1997 (first entry)
XX
XX Immunogenic type F botulinum toxin polypeptide (aa848-991).
XX
XX Botulinum toxin; neurotoxin; BoBt/F; immunogen; vaccine; botulism.
XX
XX Clostridium botulinum type F strain Langeland.
XX
XX W09641881-A1.
XX
XX 27-DEC-1996.
XX
XX 12-JUN-1996; 96MO-GB01409.
XX
XX 12-JUN-1995; 95GB-0011909.
XX
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
XX Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX WPT; 1997-065467/06.
XX
XX Immunogenic type F botulinum toxin polypeptide(s) - allows
XX recombinant vaccine prodn.
XX
XX Claim 5; Page 17-18; 37pp; English.
XX
XX Novel polypeptides (AAW09014-17) respectively comprise amino acids
```

CC 848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a  
 CC type F botulinum neurotoxin (BoNT/F). They lack the L chain and  
 CC HN epitopes necessary for metalloprotease activity and toxin  
 CC internalisation. They are free of botulinum toxin activity but can  
 CC induce protective immunity to a type F botulinum toxin, making them  
 CC useful for vaccine prodn. Recombinant polypeptides can be  
 CC produced in transformed host cells, esp. as fusion proteins, e.g.  
 CC with maltose binding protein to facilitate purification.

XX Sequence 144 AA;

Query Match 100.0%; Score 144; DB 18; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-141;

Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTNDKILILFENKLYKKIKDNIIDMRKENNKFIDISGYSNLSINGDYIYSTNRQF 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 sytdnklllyfnklykkikdksnslidmryennkfdisgysnsinsgdylystnrqf 60  
 OY 61 GIYSSKPESEVNIQONDIIVNGRYQNSISFWRIIPKYFNKVNLNNEYTIIDCIRNNSG 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 glyskspeevniaqndliygryqnfisfwrilpkylfnkvnlnneytlidcitrnnsq 120  
 OY 121 WKISLNVKIIITWTLODTAGNOKL 144  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 wkislrvnkliwltlqdtagnqkl 144

RESULT 2

AAW09014  
 ID AAW09014 standard; Protein: 431 AA.

XX AAW09014;

XX 31-MAR-1997 (first entry)

DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).

KW Botulinum toxin; neurotoxin; BoNT/F; immunogen; vaccine; botulism.

OS Clostridium botulinum type F strain Langeland.

XX W09641881-A1.

XX 27-DEC-1996.

PF 12-JUN-1996; 96WO-GB01409.

XX 12-JUN-1995; 95GB-0011909.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;

DR WPI: 1997-065467/06.

DR N-PSDB; AAT48100.

PT Immunogenic type F botulinum toxin polypeptide(s) - allows

XX recombinant vaccine prodn.

XX Claim 5; Page 16-17; 37pp; English.

XX A polypeptide (AAW09014) comprises the heavy chain (amino acids  
 CC 848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be  
 CC produced using a synthetic gene (AAT48101) based on the natural  
 CC gene sequence (AAT48100) for the heavy chain. The polypeptides and  
 CC its fragments (see also AAW09015-17) lack the light chain and HN  
 CC epitopes necessary for metalloprotease activity and toxin  
 CC internalisation. They are free of botulinum toxin activity but can  
 CC induce protective immunity to a type F botulinum toxin, making them  
 CC useful for vaccine prodn. Recombinant polypeptides can be  
 CC produced in transformed host cells, esp. as fusion proteins, e.g.

CC with maltose binding protein to facilitate purification.

XX Sequence 431 AA;

Query Match 100.0%; Score 144; DB 18; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-141;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTNDKILILFENKLYKKIKDNIIDMRKENNKFIDISGYSNLSINGDYIYSTNRQF 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 sytdnklllyfnklykkikdksnslidmryennkfdisgysnsinsgdylystnrqf 60  
 OY 61 GIYSSKPESEVNIQONDIIVNGRYQNSISFWRIIPKYFNKVNLNNEYTIIDCIRNNSG 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 glyskspeevniaqndliygryqnfisfwrilpkylfnkvnlnneytlidcitrnnsq 120  
 OY 121 WKISLNVKIIITWTLODTAGNOKL 144  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 wkislrvnkliwltlqdtagnqkl 144

RESULT 3

AAW04096  
 ID AAW04096 standard; Protein: 432 AA.

XX AAW04096;

XX 11-APR-2001 (first entry)

DE Botulinum toxin heavy chain C-terminal sequence (serotype F).

KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;

KW recombinant vector; antigen; immune response; vaccine; bacterium;

XX infection.

OS Synthetic.

XX Clostridium botulinum.

XX W0200067700-A2.

XX 16-NOV-2000.

PD 12-MAY-2000; 2000WO-US12890.

XX 12-MAY-1999; 99US-0133865.

PR 12-MAY-1999; 99US-0133866.

PR 12-MAY-1999; 99US-0133867.

PR 12-MAY-1999; 99US-0133868.

PR 12-MAY-1999; 99US-0133869.

PR 12-MAY-1999; 99US-0133870.

PR 29-JUL-1999; 99US-0146192.

XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

XX WPI: 2001-016048/02.

XX N-PSDB; AAA54490.

PT New nucleic acids encoding the carboxy- or amino-terminal portions of

XX the heavy chain of botulinum neurotoxin of serotype A-G, useful as

XX vaccine against botulism

XX Claim 3; Fig 9b; 73pp; English.

XX Botulism neurotoxins are translated as a single 150 kDa polypeptide

XX chain and then posttranslationally nicked, forming a dimeric

XX consisting of a 100 kDa heavy chain and a 50 kDa light chain which

XX remain linked by a disulfide bond. Nucleic acids encoding the

XX carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy

XX chain of botulinum neurotoxin (BoNT) can be used in recombinant

XX expression vectors and expressed in transformed cells to produce

CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organism such as *Escherichia coli* or *Pichia pastoris*. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from *Clostridium botulinum* serotypes A-G.

SQ Sequence 432 AA;

Query Match 100.0%; Score 144; DB 22; Length 432;  
Best Local Similarity 100.0%; Pred. No. 7,6e-141;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNKDKILILYFNKLYKKIKDINSILDMRYENKFEIDISGSGNSISINGDYIYSTNRNPF 60  
Db 2 sytnckillilfyfknklykkikdinsildmryenkfidsygsnslsingdylystnrngf 61

QY 61 GIYSSKPESEVNIAQNNDIYNGRYONFSIFWVRIPKYPKNKYNLNNEETIIDCIKNNNSG 120  
Db 62 glyskspeevnlaqndilnyngryonfsisfwrtpkynkynlnneytlidcinnngs 121

QY 121 WKISLANKKIITWTIADTRAGNNOKL 144  
Db 122 wkislanykklitwldttagngkl 145

RESULT 4  
AAE04103 standard; Protein: 432 AA.

AC AAE04103;  
DT 11-APR-2001 (first entry)

DE Botulism toxin heavy chain C-terminal sequence (serotype F).

KW Botulism: toxin; neurotoxin; heavy chain; recombinant expression;  
KW recombinant vector; antigen; immune response; vaccine; bacterium;  
KW infection.

OS Synthetic.  
OS *Clostridium botulinum*.

XX WO200067700-A2.  
XX  
XX 16-NOV-2000.  
XX  
XX 12-MAY-2000; 2000WO-US12890.  
XX  
XX 12-MAY-1999; 99US-0133865.  
XX 12-MAY-1999; 99US-0133866.  
XX 12-MAY-1999; 99US-0133867.  
XX 12-MAY-1999; 99US-0133868.  
XX 12-MAY-1999; 99US-0133869.  
XX 12-MAY-1999; 99US-0133870.  
XX 29-JUL-1999; 99US-0146192.

PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
XX  
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
XX  
XX WPI; 2001-016048/02.  
XX N-PSDB; AAA54499.  
XX  
XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
XX the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
XX vaccine against botulism  
XX  
XX Disclosure; Fig 18b; 73pp; English.

XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked, forming a dichain  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organism such as *Escherichia coli* or *Pichia pastoris*. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from *Clostridium botulinum* serotypes A-G.

SQ Sequence 432 AA;

Query Match 100.0%; Score 144; DB 22; Length 432;  
Best Local Similarity 100.0%; Pred. No. 7,6e-141;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTNKDKILILYFNKLYKKIKDINSILDMRYENKFEIDISGSGNSISINGDYIYSTNRNPF 60  
Db 2 sytnckillilfyfknklykkikdinsildmryenkfidsygsnslsingdylystnrngf 61

QY 61 GIYSSKPESEVNIAQNNDIYNGRYONFSIFWVRIPKYPKNKYNLNNEETIIDCIKNNNSG 120  
Db 62 glyskspeevnlaqndilnyngryonfsisfwrtpkynkynlnneytlidcinnngs 121

QY 121 WKISLANKKIITWTIADTRAGNNOKL 144  
Db 122 wkislanykklitwldttagngkl 145

RESULT 5  
AAE07894 standard; Protein: 645 AA.

AC AAE07894;  
DT 01-NOV-2001 (first entry)

DE Modified clostridial heavy chain fragment #1.

XX  
XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
KW diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.

OS Chimeric - *Corynebacterium diphtheriae*.  
OS Chimeric - *Clostridium botulinum*.

XX WO200158936-A2.  
XX  
XX 16-AUG-2001.  
XX  
XX 04-DEC-2000; 2000WO-GB04644.  
XX  
XX 02-DEC-1999; 99GB-0028530.  
XX 07-APR-2000; 2000GB-0008658.  
XX  
XX (MIGR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
XX Shone CC, Sutton JM, Silman N;  
XX  
XX WPI; 2001-514643/56.  
XX  
XX New non toxic polypeptide for delivery of a therapeutic agent for the  
XX treatment of a CNS disorder comprising a binding domain that  
XX translocates the therapeutic agent into the neuronal cells -  
XX

```
XX PS Example 2; Page 44; 50pp; English.
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain of botulinum neurotoxin type F
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 645 AA;

Query Match 100.0%; Score 144; DB 22; Length 645;
Best Local Similarity 100.0%; Pred. No. 1.1e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTNDKILIIYFNKLYKKIKDNIIDMRKRNKRFIDISGYSNINSINGDVIYSTNRNOF 60
DB 215 sytdnkiliiyfnklykkikdnlidmrkrfidisgysnsinsingdviystnrngf 274
OY 61 GIYSKSPSEVNIAQNNDIYNGRYQNFISFWVRIPRYFNKNVNNNEYTIIDCIRNNNSG 120
DB 275 giyskspsevnlaqndliygryqnfisfwvripkyfnknvnnneytliidcirmnsg 334
OY 121 WKISLANYNKIIWTLODPAGNOKL 144
DB 335 wkislanynkliwtlqdgagmqkl 358

RESULT 6
ID AAE07898 standard; Protein; 660 AA.
AC AAE07898;
DT 01-NOV-2001 (first entry)
DE Modified clostridial heavy chain fragment #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW diphtheria neurotoxin; tetanus neurotoxin; TeNT;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium tetani.
OS Chimeric - Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
```

```
XX PS New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
XX PS Example 2; Page 46; 50pp; English.
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain which is a hybrid of botulinum
CC neurotoxin type F (BoNT/F) and tetanus neurotoxin (TeNT) domain II with
CC translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 660 AA;

Query Match 100.0%; Score 144; DB 22; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.1e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYTNDKILIIYFNKLYKKIKDNIIDMRKRNKRFIDISGYSNINSINGDVIYSTNRNOF 60
DB 215 sytdnkiliiyfnklykkikdnlidmrkrfidisgysnsinsingdviystnrngf 274
OY 61 GIYSKSPSEVNIAQNNDIYNGRYQNFISFWVRIPRYFNKNVNNNEYTIIDCIRNNNSG 120
DB 275 giyskspsevnlaqndliygryqnfisfwvripkyfnknvnnneytliidcirmnsg 334
OY 121 WKISLANYNKIIWTLODPAGNOKL 144
DB 335 wkislanynkliwtlqdgagmqkl 358

RESULT 7
ID AAE07893 standard; Protein; 685 AA.
AC AAE07893;
DT 01-NOV-2001 (first entry)
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Influenza virus.
OS Chimeric - Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
```

PA (MCCR-) MICROBIOLOGICAL RES AUTHORITY.  
 XX  
 PI Shone CC, Sutton JM, Silman N;  
 XX  
 DR WPI: 2001-514643/56.  
 XX  
 PT New non toxic polypeptide for delivery of a therapeutic agent for the  
 PT treatment of a CNS disorder comprising a binding domain that  
 PT translocates the therapeutic agent into the neuronal cells -  
 XX  
 PS Example 9; Page 43; 50pp; English.  
 XX  
 CC The invention relates to a non toxic polypeptide, for delivery of a  
 CC therapeutic agent to a neuronal cell, which comprises a binding domain  
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
 CC as Hc) that binds to the neuronal cell and a translocation domain (amino  
 CC terminal half of HC, designated as HN), that translocates the therapeutic  
 CC agent into the neuronal cell, where the translocation domain is not a HN  
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
 CC useful for the treatment of a disease state associated with neuronal  
 CC cells. The polypeptide constructs are useful for delivering therapeutic  
 CC substances to neuronal cells. They are useful to treat disorders of the  
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
 CC and infection. They are also useful in gene therapy. The present sequence  
 CC is modified clostridial heavy chain-superoxide dismutase conjugate. This  
 CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from  
 CC *Bacillus stearothermophilus*, linker that can be cleaved by factor Xa,  
 CC translocation peptide from influenza virus and a neuronal cell-specific  
 CC binding domain from botulinum neurotoxin type F (BONT/F).  
 XX  
 SQ Sequence 685 AA:  
 Query Match 100.0%; Score 144; DB 22; Length 685;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-140;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SYTNDKTLLILYFNKLYKKIKDSTLDMRYENKKFTDISGYSNISTINGDYIYSTNRNQF 60  
 Db 255 sytnckllilfyfnyklykkikdkslldmryenkkftdisgysnistsingdyystnngf 314  
 QY 61 GYSSKRPSEVNIAQNNDIYNGRYONESISFWVRIPKRYFNKYNLNNEYTTIDCIRNNNSG 120  
 Db 315 gyskspsevnlaqndilnygrynsisfwrpkyfnkynlnneytldcitrnnsg 374  
 QY 121 WKISLNVNKKITWTLODTAGNNOKL 144  
 Db 375 wkislwnkkitwtltdtagnngkl 398  
 RESULT 8  
 AAE07890  
 ID AAE07890 standard; Protein: 862 AA.  
 AC AAE07890;  
 XX  
 DT 01-NOV-2001 (first entry)  
 XX  
 DE Modified clostridial heavy chain-superoxide dismutase conjugate #2.  
 XX  
 KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
 KM tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
 KM superoxide dismutase; SOD; diphtheria neurotoxin;  
 KM botulinum neurotoxin type F; BONT/F.  
 XX  
 OS Chimeric - *Bacillus stearothermophilus*.  
 OS Chimeric - *Corynebacterium diphtheriae*.  
 OS Chimeric - Clostridium botulinum.  
 OS Chimeric - Synthetic.  
 XX  
 PN WO200158936-A2.

PD 16-AUG-2001.  
 XX  
 PF 04-DEC-2000; 2000WO-GB04644.  
 XX  
 PR 02-DEC-1999; 99GB-0028530.  
 PR 07-APR-2000; 2000GB-0008658.  
 XX  
 PA (MCCR-) MICROBIOLOGICAL RES AUTHORITY.  
 XX  
 PI Shone CC, Sutton JM, Silman N;  
 XX  
 DR WPI: 2001-514643/56.  
 XX  
 PT New non toxic polypeptide for delivery of a therapeutic agent for the  
 PT treatment of a CNS disorder comprising a binding domain that  
 PT translocates the therapeutic agent into the neuronal cells -  
 XX  
 PS Example 9; Page 40; 50pp; English.  
 XX  
 CC The invention relates to a non toxic polypeptide, for delivery of a  
 CC therapeutic agent to a neuronal cell, which comprises a binding domain  
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
 CC as Hc) that binds to the neuronal cell and a translocation domain (amino  
 CC terminal half of HC, designated as HN), that translocates the therapeutic  
 CC agent into the neuronal cell, where the translocation domain is not a HN  
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
 CC useful for the treatment of a disease state associated with neuronal  
 CC cells. The polypeptide constructs are useful for delivering therapeutic  
 CC substances to neuronal cells. They are useful to treat disorders of the  
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
 CC and infection. They are also useful in gene therapy. The present sequence  
 CC is modified clostridial heavy chain-superoxide dismutase conjugate.  
 CC This conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from  
 CC *Bacillus stearothermophilus*, linker that can be cleaved by factor Xa,  
 CC translocation domain from diphtheria neurotoxin and a neuronal cell-  
 CC specific binding domain from botulinum neurotoxin type F (BONT/F).  
 XX  
 SQ Sequence 862 AA:  
 Query Match 100.0%; Score 144; DB 22; Length 862;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-140;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SYTNDKTLLILYFNKLYKKIKDSTLDMRYENKKFTDISGYSNISTINGDYIYSTNRNQF 60  
 Db 432 sytnckllilfyfnyklykkikdkslldmryenkkftdisgysnistsingdyystnngf 491  
 QY 61 GYSSKRPSEVNIAQNNDIYNGRYONESISFWVRIPKRYFNKYNLNNEYTTIDCIRNNNSG 120  
 Db 492 gyskspsevnlaqndilnygrynsisfwrpkyfnkynlnneytldcitrnnsg 551  
 QY 121 WKISLNVNKKITWTLODTAGNNOKL 144  
 Db 552 wkislwnkkitwtltdtagnngkl 575  
 RESULT 9  
 AAE07892  
 ID AAE07892 standard; Protein: 887 AA.  
 AC AAE07892;  
 XX  
 DT 01-NOV-2001 (first entry)  
 XX  
 DE Modified clostridial heavy chain-superoxide dismutase conjugate #4.  
 XX  
 KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
 KM tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
 KM superoxide dismutase; SOD; diphtheria neurotoxin; human;  
 KM botulinum neurotoxin type F; BONT/F.  
 XX

OS Chimeric - Homo sapiens.  
 OS Chimeric - Bacillus stearothermophilus.  
 OS Chimeric - Corynebacterium diphtheriae.  
 OS Chimeric - Clostridium botulinum.  
 OS Chimeric - Synthetic.  
 XX  
 XX WO200158936-A2.  
 XX  
 XX 16-AUG-2001.  
 PD  
 XX  
 PF 04-DEC-2000; 2000WO-GB04644.  
 XX  
 XX 02-DEC-1999; 99GB-0028530.  
 PR 07-APR-2000; 2000GB-0008658.  
 XX  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 PI Shone CC, Sutton JM, Silman N;  
 XX  
 DR WPI; 2001-514643/56.  
 XX  
 XX  
 PT New non toxic polypeptide for delivery of a therapeutic agent for the  
 PT treatment of a CNS disorder comprising a binding domain that  
 PT translocates the therapeutic agent into the neuronal cells -  
 XX  
 XX Example 9; Page 42; 50pp; English.  
 PS  
 XX  
 CC The invention relates to a non toxic polypeptide, for delivery of a  
 CC therapeutic agent to a neuronal cell, which comprises a binding domain  
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
 CC as Hc) that binds to the neuronal cell and a translocation domain (amino  
 CC terminal half of HC, designated as HN), that translocates the therapeutic  
 CC agent into the neuronal cell, where the translocation domain is not a HN  
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
 CC useful for the treatment of a disease state associated with neuronal  
 CC cells. The polypeptide constructs are useful for delivering therapeutic  
 CC substances to neuronal cells. They are useful to treat disorders of the  
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
 CC and infection. They are also useful in gene therapy. The present sequence  
 CC is modified clostridial heavy chain-superoxide dismutase conjugate.  
 CC This conjugate comprises a mitochondrial leader sequence from human  
 CC Mn-superoxide dismutase (MnSOD). MnSOD from Bacillus stearothermophilus,  
 CC linker that can be cleaved by thrombin, translocation domain from  
 CC diphtheria neurotoxin and a neuronal cell-specific binding domain from  
 CC botulinum neurotoxin type F (BoNT/F).  
 CC  
 XX Sequence 887 AA;  
 SQ

Query Match 100.0%; Score 144; DB 22; Length 887;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-140;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYNDKLLILYFNKLYKKIKDNLDMRYENKRFIDISGYSNLSINGDYIYSTNNQF 60  
 DB 457 syndkllillyfnklykkikdnlldmryenkrfidisgysnlsingdyiystnnqf 516  
 QY 61 GYSSKSEVNIAQNDIIYNGRYONFSIFWVRIPKYPKVNLNNEYTIIDCIRNNNSG 120  
 DB 517 gyskspsevnlaqndiilyngryonfsisfwiripkyfknvlnneytliidcirmnsg 576  
 QY 121 WKISLNTNKKIITWLTQDTAGNNQKL 144  
 DB 577 wkislntnkkiiwtlqdtagnnqkl 600

RESULT 10  
 AAE07901  
 ID AAE07901 standard; protein; 1032 AA.  
 XX  
 AC AAE07901;  
 XX

DT 01-NOV-2001 (first entry)  
 XX  
 DE C. botulinum C2 translocation domain with BoNT/F-binding domain #2.  
 XX  
 KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
 KW tumour; infection; neurodegenerative disease; gene therapy;  
 KW botulinum neurotoxin type F; BoNT/F.  
 XX  
 OS Clostridium botulinum.  
 XX  
 XX WO200158936-A2.  
 PN  
 XX  
 XX 16-AUG-2001.  
 PD  
 XX  
 PF 04-DEC-2000; 2000WO-GB04644.  
 XX  
 XX 02-DEC-1999; 99GB-0028530.  
 PR 07-APR-2000; 2000GB-0008658.  
 XX  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 PI Shone CC, Sutton JM, Silman N;  
 XX  
 DR WPI; 2001-514643/56.  
 XX  
 XX  
 PT New non toxic polypeptide for delivery of a therapeutic agent for the  
 PT treatment of a CNS disorder comprising a binding domain that  
 PT translocates the therapeutic agent into the neuronal cells -  
 XX  
 XX Example 2; Page 48; 50pp; English.  
 PS  
 XX  
 CC The invention relates to a non toxic polypeptide, for delivery of a  
 CC therapeutic agent to a neuronal cell, which comprises a binding domain  
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
 CC as Hc) that binds to the neuronal cell and a translocation domain (amino  
 CC terminal half of HC, designated as HN), that translocates the therapeutic  
 CC agent into the neuronal cell, where the translocation domain is not a HN  
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
 CC useful for the treatment of a disease state associated with neuronal  
 CC cells. The polypeptide constructs are useful for delivering therapeutic  
 CC substances to neuronal cells. They are useful to treat disorders of the  
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
 CC and infection. They are also useful in gene therapy. The present sequence  
 CC is C. botulinum C2 enterotoxin translocation domain with botulinum  
 CC neurotoxin type F (BoNT/F) binding domain used in the exemplification of  
 CC the invention.  
 CC  
 XX Sequence 1032 AA;  
 SQ

Query Match 100.0%; Score 144; DB 22; Length 1032;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-140;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYNDKLLILYFNKLYKKIKDNLDMRYENKRFIDISGYSNLSINGDYIYSTNNQF 60  
 DB 602 syndkllillyfnklykkikdnlldmryenkrfidisgysnlsingdyiystnnqf 661  
 QY 61 GYSSKSEVNIAQNDIIYNGRYONFSIFWVRIPKYPKVNLNNEYTIIDCIRNNNSG 120  
 DB 662 gyskspsevnlaqndiilyngryonfsisfwiripkyfknvlnneytliidcirmnsg 721  
 QY 121 WKISLNTNKKIITWLTQDTAGNNQKL 144  
 DB 722 wkislntnkkiiwtlqdtagnnqkl 745

RESULT 11  
 AAY93309  
 ID AAY93309 standard; protein; 1059 AA.  
 XX  
 AC AAY93309;  
 XX



```
XX 04-SEP-2000 (first entry)
DT
XX
XX A manganese superoxide dismutase (Mn-SOD) construct.
DE
XX
XX Manganese superoxide dismutase: Mn-SOD; SOD: neuronal cell;
KM neuronal cell targeting component: NCTC; neuronal disease:
KM oxidative stress: ischemic stroke; trauma: Parkinson's disease;
KM Huntington's disease; motor neurone disease;
KM botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Bacillus stearothermophilus.
OS Clostridium botulinum.
XX
XX WO200028041-A1.
XX
XX 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-GB03699.
XX
XX 05-NOV-1998; 98GB-0024282.
XX
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
XX Shone CC, Sutton JM, Hallis B, Silman N;
XX WPI; 2000-376553/32.
XX
XX Novel composition, comprising superoxide dismutase linked by a
PT cleavable linker to a neuronal cell targeting component useful for
PT delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
XX Disclosure; Page 48-51; 65pp; English.
XX
XX The present sequence represents a construct of the invention, comprising
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
CC can be cleaved by thrombin, and a heavy chain derived from botulinum
CC neurotoxin serotype F. The specification describes a composition for
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC This component has a domain that binds to a neuronal cell and a
CC domain that translocates the SOD of the composition into the neuronal
CC cell. After translocation, the linker is cleaved to release the SOD.
CC The composition is useful for treating neuronal diseases caused or
CC augmented by oxidative stress, such as ischemic stroke, trauma,
CC Parkinson's disease, Huntington's disease and motor neurone diseases.
CC
XX
XX Sequence 1059 AA:
SQ
Query Match 100.0%; Score 144; DB 21; Length 1059;
Best Local Similarity 100.0%; Pred. No. 1.7e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYTNDKILLYFNKLYKRIKDNSTIDMYENKPFIDISGYSNINSGDYIYSTNRNF 60
DB 629 sytdnklllyfnyklykikdnslidmyenkfidisgysninsngdyiystrnngf 688
QY 61 GYSSKSEVNIAQNNDIYNGRNFSEISFWRIKRYKYNLNNEETITDCIRNNNSG 120
DB 689 gylsskspsevnlaqndiylngrynfseisfwrlpkyfknvnlneeytldictirnnsg 748
QY 121 WKISLWYKRIIWTLODTAGNOKL 144
DB 749 wkislwnkriiwtlqdtagnokl 772
RESULT 12
ID AAY93312 standard; protein: 1084 AA.
XX
XX AAY93312;
```

```
XX 04-SEP-2000 (first entry)
DT
XX
XX A manganese superoxide dismutase (Mn-SOD) construct.
DE
XX
XX Manganese superoxide dismutase: Mn-SOD; SOD: neuronal cell;
KM neuronal cell targeting component: NCTC; neuronal disease:
KM oxidative stress: ischemic stroke; trauma: Parkinson's disease;
KM Huntington's disease; motor neurone disease;
KM botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Homo sapiens.
OS Bacillus stearothermophilus.
OS Clostridium botulinum.
XX
XX WO200028041-A1.
XX
XX 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-GB03699.
XX
XX 05-NOV-1998; 98GB-0024282.
XX
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
XX Shone CC, Sutton JM, Hallis B, Silman N;
XX WPI; 2000-376553/32.
XX
XX Novel composition, comprising superoxide dismutase linked by a
PT cleavable linker to a neuronal cell targeting component useful for
PT delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
XX Disclosure; Page 57-60; 65pp; English.
XX
XX The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide
CC dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker
CC that can be cleaved by thrombin, and a heavy chain derived from
CC botulinum neurotoxin serotype F. The specification describes a
CC composition for delivery of SOD to neuronal cells. The composition
CC comprises SOD linked, by a cleavable linker, to a neuronal cell
CC targeting component (NCTC). This component has a domain that binds
CC to a neuronal cell and a domain that translocates the SOD of the
CC composition into the neuronal cell. After translocation, the linker
CC is cleaved to release the SOD. The composition is useful for treating
CC neuronal diseases caused or augmented by oxidative stress, such as
CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases.
CC
XX
XX Sequence 1084 AA:
SQ
Query Match 100.0%; Score 144; DB 21; Length 1084;
Best Local Similarity 100.0%; Pred. No. 1.8e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYTNDKILLYFNKLYKRIKDNSTIDMYENKPFIDISGYSNINSGDYIYSTNRNF 60
DB 654 sytdnklllyfnyklykikdnslidmyenkfidisgysninsngdyiystrnngf 713
QY 61 GYSSKSEVNIAQNNDIYNGRNFSEISFWRIKRYKYNLNNEETITDCIRNNNSG 120
DB 714 gylsskspsevnlaqndiylngrynfseisfwrlpkyfknvnlneeytldictirnnsg 773
QY 121 WKISLWYKRIIWTLODTAGNOKL 144
DB 774 wkislwnkriiwtlqdtagnokl 797
RESULT 13
ID AAE07900
```

ID AAE07900 standard; Protein; 1092 AA.  
XX AAE07900;  
AC AAE07900;  
XX  
XX  
DT 01-NOV-2001 (first entry)  
XX  
XX  
DE C. botulinum C2 translocation domain with BONT/F-binding domain #1.  
XX  
XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KM tumour; infection; neurodegenerative disease; gene therapy;  
KM botulinum neurotoxin type F; BONT/F.  
XX  
XX Clostridium botulinum.  
OS  
XX WO200158936-A2.  
PN  
XX 16-AUG-2001.  
PD  
XX 04-DEC-2000; 2000WO-GB04644.  
PF  
XX 02-DEC-1999; 99GB-0028530.  
PR 07-APR-2000; 2000GB-0008658.  
XX  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
PA  
XX Shone CC, Sutton JM, Silman N;  
PI  
XX WPI; 2001-514643/56.  
DR  
XX  
XX New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -  
PT  
XX  
PS Example 2; Page 47; 50pp; English.  
XX  
XX The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as Hc) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial neurotoxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is C. botulinum C2 enterotoxin translocation domain with botulinum  
CC neurotoxin type F (BONT/F) binding domain used in the exemplification of  
CC the invention.  
XX  
XX Sequence 1092 AA;  
SQ  
Query Match 100.0%; Score 144; DB 22; Length 1092;  
Best Local Similarity 100.0%; Pred. No. 1.8e-140;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYTNDKILILYFNKLYKKIKDNTSLDMRYENKRFIDISGGSNLSINGDVIYSTNRNOF 60  
Db 662 sytdkllillyfnklykkikdntslldmryennkfidisggsnlsingdviystnrgf 721  
QY 61 GIYSSKPESEVIAQNDIINYGRYONFSISFWVRIPEYFNKVNINNETIICIRNNNSG 120  
Db 722 giyskpesevniagndiinygrysfsisfwiripkyfknvlnneytldcrlmnsg 781  
QY 121 WKISLWNNKIITWLTQTFAGNNOKL 144  
Db 782 wkislwnnkiitwltqtfagnnqkl 805  
RESULT 14

AAV77138  
ID AAV77138 standard; Protein; 432 AA.  
XX AAV77138;  
AC AAV77138;  
XX  
XX  
DT 08-MAY-2000 (first entry)  
XX  
XX  
DE Synthetic botulinum neurotoxin serotype F (BONTF) C-terminal fragment.  
XX  
XX Botulinum neurotoxin; heavy chain; BONT; serotype F;  
KM C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
KM VEE; botulism; vaccine; diagnosis; drug screening.  
XX  
XX Clostridium botulinum.  
OS  
XX Synthetic.  
XX  
XX WO200002524-A2.  
PN  
XX 20-JAN-2000.  
PD  
XX  
XX 09-JUL-1999; 99WO-US15570.  
PF  
XX  
XX 10-JUL-1998; 98US-0092416.  
PR 12-MAY-1999; 99US-0133870.  
XX  
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
PA  
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
PI  
XX WPI; 2000-160827/14.  
DR  
XX N-PsDB; AA287216.  
XX  
XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
PT toxin serotypes A-G, is used for inducing an immune response against  
PT botulinum -  
PT  
XX  
XX Claim 27; Page -; 54pp; English.  
PS  
XX  
XX The invention relates to novel vaccines that induce a protective immune  
CC response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F  
CC and G (BONTA-BONTG). The vaccine of the invention is novel recombinant  
CC DNA construct comprising a vector, and at least one nucleic acid  
CC fragment comprising a C-terminal heavy chain fragment (HC) from BONT  
CC serotypes A-G. In preferred embodiments of the invention, the vector is  
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
CC this vector results in the production of large amounts of a protein  
CC encoded by a sequence cloned into the replicon. The constructs are used  
CC to produce vaccines against botulism. The proteins can also be used as  
CC diagnostic tools for the diagnosis of botulism. The transformed host  
CC cells can be used to analyse the effectiveness of drugs and agents which  
CC inhibit toxin effects. The vaccine currently used against botulism is  
CC dangerous and expensive to produce, and contains formalin, which is very  
CC painful for the recipient. Also, the vaccine is incomplete, in that only  
CC 5 of the 7 serotypes are represented in the formulation. The novel  
CC vaccine overcomes these problems, as it is easily purified, and  
CC available in large quantities. It is also expressed in the lymph nodes  
CC for a better immune response. Sequences AAV77134-Y77139 represent  
CC synthetic BONT Hc fragments used in the present invention. The DNA  
CC encoding these sequences had been optimised for codon usage for  
CC expression in yeast. Note: This sequence is not given in the  
CC specification, but is decoded from the BONTF Hc DNA sequence given on  
CC pages 45-46.  
XX  
XX Sequence 432 AA;  
SQ  
Query Match 61.1%; Score 88; DB 21; Length 432;  
Best Local Similarity 100.0%; Pred. No. 8.1e-83;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYTNDKILILYFNKLYKKIKDNTSLDMRYENKRFIDISGGSNLSINGDVIYSTNRNOF 60  
Db 2 sytdkllillyfnklykkikdntslldmryennkfidisggsnlsingdviystnrgf 61

OY 61 GIYSSKPSSEVNIQONNDIYNGRQNES 88  
 ||||||||||||||||||||||||||||  
 Db 62 gIYSSKPSSEVNIQONNDIYNGRYGNFS 89

Search completed: August 15, 2002, 11:12:26  
 Job time: 318 sec

## RESULT 15

AAW68399

ID AAW68399 standard; Protein; 448 AA.

XX AAW68399;

XX 07-DEC-1998 (first entry)

XX Clostridium botulinum type F toxin C fragment.

XX Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;

KW botulism; BotF.

XX Clostridium botulinum serotype F strain 202F (ATCC 23387).

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..21 /note= "N-terminal His tag"

XX WO9808540-A1.

XX 05-MAR-1998.

XX 28-AUG-1997; 97WO-US15394.

XX 28-AUG-1996; 96US-0704159.

XX (OPHI-) OPHIDIAN PHARM INC.

XX Thalley BS, Williams JA;

XX WPI: 1998-230234/20.

XX N-PSDB; AAV30593.

XX Host cell containing recombinant expression vector encoding  
 PT Clostridium botulinum type B or E toxin - useful to treat humans  
 PT and other animals at risk of intoxication with clostridial toxin

XX Example 48; Page 364-365; 428pp; English.

XX This is the amino acid sequence of the histidine-tagged C fragment  
 CC of Clostridium botulinum (202F strain) type F neurotoxin, encoded  
 CC by a DNA sequence (see AAV30593) in plasmid pETH10. This vector  
 CC can be used to express BotC soluble C fragment in Escherichia  
 CC coli host cells, with the recombinant C fragment being purified on  
 CC an affinity column. The invention relates to recombinant proteins  
 CC derived from C. botulinum toxins, especially type B and type E  
 CC soluble recombinant proteins free of significant endotoxin  
 CC contamination. Preferred hosts for production of recombinant  
 CC proteins are E. coli, insect cells and yeast cells. The  
 CC recombinant toxins are used as immunogens for the production of  
 CC vaccines and antitoxins that are useful in the treatment of humans  
 CC and animals at risk of intoxication with clostridial toxin.

XX Sequence 448 AA;

Query Match 18.1%; Score 26; DB 19; Length 448;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 SIIDMRYENKRFIDSGYSGNSISNG 48  
 ||||||||||||||||||||||||||||  
 Db 43 SIIDMRYENKRFIDSGYSGNSISNG 68

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:37 ; Search time 24.69 Seconds  
(without alignments) 225.825 Million cell updates/sec

Title: US-08-981-087A-2

Perfect score: 144  
Sequence: 1 STYNDKILILYFNKIKRKIK.....LNYKKIITLQDTAGNCKL 144

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	18.1	1274	1	BXF_CLOBO
2	11	7.6	1250	1	BXE_CLOBO
3	11	7.6	1250	1	BXE_CLOBO
4	9	6.2	1285	1	BXA2_CLOBO
5	8	5.6	1295	1	BXA_CLOBO
6	7	4.9	1449	1	MRFE_RICPR
7	7	4.9	501	1	VGIC_HSYMB
8	7	4.9	501	1	VGIC_HSYMB
9	7	4.9	501	1	VGIC_HSYMB
10	7	4.9	501	1	VGIC_HSYMB
11	7	4.9	505	1	VGIC_HSYMB
12	7	4.9	511	1	RRB1_YEAST
13	7	4.9	2292	1	POIG_EMCVB
14	7	4.9	2292	1	POIG_EMCVB
15	6	4.2	79	1	CATR_HUMAN
16	6	4.2	112	1	YF88_MERJA
17	6	4.2	117	1	RR20_ASTLO
18	6	4.2	134	1	GILZ_HUMAN
19	6	4.2	134	1	GILZ_HUMAN
20	6	4.2	140	1	LYC_AMOGA
21	6	4.2	153	1	COX2_COMRU
22	6	4.2	171	1	BB19_RABIT
23	6	4.2	176	1	CCTR_MACMU
24	6	4.2	180	1	CCTR_CAYPO
25	6	4.2	180	1	Y426_MERJA
26	6	4.2	193	1	Y262_HERY
27	6	4.2	193	1	Y262_HERY
28	6	4.2	196	1	Y264_BACSU
29	6	4.2	204	1	Y264_BACSU
30	6	4.2	205	1	Y264_BACSU
31	6	4.2	211	1	Y264_BACSU
32	6	4.2	214	1	Y264_BACSU
33	6	4.2	214	1	Y264_BACSU

34	6	4.2	214	1	CYB_BOTSC	P2849 bothriechis
35	6	4.2	217	1	EXPI_ERMCA	P33882 erwina car
36	6	4.2	217	1	YE9H_SCHPO	O13777 schizosach
37	6	4.2	218	1	Y010_MYCGE	P47256 mycoplasma
38	6	4.2	226	1	TDX1_CAEL	O21824 caenorhabd1
39	6	4.2	243	1	NGF_BUNMU	P34128 bungarus mu
40	6	4.2	245	1	PLSC_SALTY	P26974 salmoneilla
41	6	4.2	247	1	Y276_BUCAT	P57564 buchnera ap
42	6	4.2	256	1	HYPA_HYPLI	P55587 hypoderma 1
43	6	4.2	257	1	TRP1_GIALA	P36186 giardia lam
44	6	4.2	257	1	TRP1_GIALA	P36187 giardia lam
45	6	4.2	278	1	NIFR_METVO	P06119 methanococc

#### ALIGNMENTS

RESULT 1	BXF_CLOBO	STANDARD:	PRT: 1274 AA.
ID	BXF_CLOBO		
AC	P30996;		
DT	01-JUL-1993 (Rel. 26, Created)		
DR	01-JUL-1993 (Rel. 26, Last sequence update)		
DI	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)		
GN	BoNT/F		
OS	Clostridium botulinum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OX	Clostridium.		
NCBI_TaxID=1491;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 23387;		
RX	MEDLINE=93012902; PubMed=1398040;		
RA	East A.K., Richardson P.T., Allaway D., Collins M.D.,		
RA	Roberts T.A., Thompson D.E.;		
RT	Sequence of the gene encoding type F neurotoxin of Clostridium		
RT	botulinum.		
RL	[2]		
RP	SEQUENCE OF 1-64 FROM N.A.		
RC	STRAIN=HOBBS FT10;		
RX	MEDLINE=94297488; PubMed=7764998;		
RA	East A.K., Collins M.D.;		
RT	Conserved structure of genes encoding components of botulinum		
RT	neurotoxin complex M and the sequence of the gene coding for the		
RT	neurotoxin component in nonproteolytic Clostridium botulinum type F.;		
RL	Curr. Microbiol. 29:69-77(1994).		
RN	[3]		
RP	SEQUENCE OF 634-1002 FROM N.A.		
RX	MEDLINE=94013372; PubMed=8408542;		
RA	Campbell K., East A.K., Collins M.D.;		
RT	"Gene probes for identification of the botulinum neurotoxin gene and		
RT	specific identification of neurotoxin types B, E, and F.;"		
RL	J. Clin. Microbiol. 31:2255-2262(1993).		
RN	[4]		
RP	IDENTIFICATION OF SUBSTRATE.		
RX	MEDLINE=94230352; PubMed=8175689;		
RA	Yamasaki S., Baumeister A., Birt T., Blas J., Link E., Cornille F.,		
RA	Rouges B., Fyfe E.M., Suedhof T.C., Jahn R., Niemann H.;		
RT	"Cleavage of members of the synaptobrevin/VAMP family by types D and		
RT	F botulinum neurotoxins and tetanus toxin.;"		
RL	J. Biol. Chem. 269:12764-12772(1994).		
CC	-1- RELEASE: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER		
CC	RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED		
CC	AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD		
CC	WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT		
CC	INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC		
CC	ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-1-LYS-59		
CC	BOND OF SYNAPTOSOMAL-1 AND -2.		
CC	-1- CATALYTIC ACTIVITY: limited hydrolysis of proteins of the		
CC	neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO		

CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M92906; AAA23263.1; -;  
 DR EMBL: S73676; AAC60475.1; -;  
 DR EMBL: X70820; CAA50151.1; -;  
 DR EMBL: X70816; CAA50147.1; -;  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -;  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOTOXILYSIN.  
 DR PRODOM: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 KM Neurotoxin; Transmembrane; Hydrolyase; Metalloprotease; Zinc  
 FT CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT CHAIN.  
 FT METAL 437 1274 BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.  
 FT ACT\_SITE 227 227 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 231 231 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 429 445 INTERCHAIN (PROBABLE).  
 FT SEQUENCE 1274 AA; 146709 MW; 5B99756A7438B921 CRC64;  
 SQ  
 Query Match 18.1%; Score 26; DB 1; Length 1274;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-19;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 SILDRENNKFDISGYGNSING 48  
 DB 869 SILDRENNKFDISGYGNSING 894  
 RESULT 2  
 BXL\_CLOBO STANDARD; PRT; 1250 AA.  
 AC 000496;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)  
 DE (Bontoxilysin E).  
 OS Clostridium botulinum.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 CC Clostridium  
 CC NCBI\_TaxID=1491;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BELUGA;  
 RX MEDLINE=92181428; PubMed=1543481;  
 RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;  
 RT "Sequences of the botulinum neurotoxin E derived from Clostridium  
 RT botulinum type E (strain Beluga) and Clostridium butyricum (strains  
 RT ATCC 43181 and ATCC 43755).";  
 RL Biochem. Biophys. Res. Commun. 183:107-113(1992).  
 RN [2]  
 RN SEQUENCE FROM N.A.

RX MEDLINE=92174922; PubMed=1541280;  
 RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;  
 RT "The complete amino acid sequence of the Clostridium botulinum type-E  
 RT neurotoxin, derived by nucleotide-sequence analysis of the encoding  
 RT gene.";  
 RL Eur. J. Biochem. 204:657-667(1992).  
 RN [3]  
 RN SEQUENCE OF 1-251 FROM N.A.  
 RX MEDLINE=90264400; PubMed=2160960;  
 RA Bliz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;  
 RT "The complete sequence of botulinum neurotoxin type A and comparison  
 RT with other clostridial neurotoxins.";  
 RL J. Biol. Chem. 265:9153-9158(1990).  
 RN [4]  
 RN SEQUENCE OF 1-13.  
 RX MEDLINE=85197963; PubMed=3888113;  
 RA Schmidt J.J., Sathymoorthy V., Dasgupta B.R.;  
 RT "Partial amino acid sequences of botulinum neurotoxins types B and  
 RT E.";  
 RL Arch. Biochem. Biophys. 238:544-548(1985).  
 RN [5]  
 RN SEQUENCE OF 419-426.  
 RX MEDLINE=90344918; PubMed=2116911;  
 RA Gimenez J.A., Dasgupta B.R.;  
 RT "Botulinum neurotoxin type E, fragmented with endoproteinase Lys-C  
 RT reveals the site trypsin nicks and homology with tetanus  
 RT neurotoxin.";  
 RL Biochimie 72:213-217(1990).  
 RN [6]  
 RN IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94063091; PubMed=8243676;  
 RA Schlavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
 RA Benfenati F., Wilson M.C., Montecucco C.;  
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 RT COOH-terminal peptide bonds.";  
 RL FEBS Lett. 335:99-103(1993).  
 RN [7]  
 RN IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94124495; PubMed=8294407;  
 RA Bliz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 RA Jahn R., Niemann H.;  
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";  
 RL J. Biol. Chem. 269:1617-1620(1994).  
 CC -1- FUNCTION: BOTULINUM TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-1-ILE-  
 CC 181 BOND IN SNAP-25.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X62089; CAA4399.1; -;  
 DR EMBL: X62683; CAA44558.1; -;  
 DR PIR: A60027; A60027.

```

DR PIR: B35294; B35294.
DR PIR: JH0257; JH0257.
DR PIR: S08575; S08575.
DR PIR: S18111; S18111.
DR PIR: S21178; S21178.
DR HSSP: P10845; 3BTA.
DR MEROPS: M27.002; .
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_Mpeptidase.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0
FT CHAIN 1 421
FT CHAIN 422 1250
FT METAL 211 211
FT ACT_SITE 212 212
FT METAL 215 215
FT DISULFID 411 425
FT CONFLICT 176 176
FT CONFLICT 197 197
FT CONFLICT 339 339
FT CONFLICT 772 772
FT CONFLICT 966 966
FT CONFLICT 1194 1194
SQ SEQUENCE 1250 AA; 143712 MW; DPFCE26DDNA41B4 CRC64;

Query Match 7.6%; Score 11; DB 1; Length 1250;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 NFSISFWVRIP 96
DB 913 NFSISFWVRIP 923

RESULT 3
BXE_CLOBO STANDARD; PRT; 1250 AA.
ID BKE_CLOBO
AC P30995;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (Bont/E)
DE (Bontoxilysin E).
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_Taxid=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43181, AND ATCC 43755;
RX MEDLINE-92181428; PubMed-1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulin neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE OF 1-251 FROM N.A.
RC STRAIN-BL6340;
RX MEDLINE-91237316; PubMed-2033376;
RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA Tokosawa N., Tashiki T., Oguma K.;
RT Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. Gen. Microbiol. 137:519-525(1991).
RN [3]
RP SEQUENCE OF 1-48.

```

```

RC STRAIN-5262;
RA Gimenez J., Foley J., Dasgupta B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum:
RT partial sequence and comparison.";
RL FASEB J. 2:A1750-A1750(1988).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62088; CAA43988.1; -.
DR EMBL: X53180; CAA37321.1; -.
DR PIR: JH0256; JH0256.
DR PIR: S16145; S16145.
DR HSSP: P10845; 3BTA.
DR MEROPS: M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_Mpeptidase.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0
FT CHAIN 1 421
FT CHAIN 422 1250
FT METAL 211 211
FT METAL 212 212
FT ACT_SITE 215 215
FT METAL 215 215
FT DISULFID 411 425
FT CONFLICT 229 229
SQ SEQUENCE 1250 AA; 143265 MW; 817B5B2C312857 CRC64;

Query Match 7.6%; Score 11; DB 1; Length 1250;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 NFSISFWVRIP 96
DB 913 NFSISFWVRIP 923

RESULT 4
BXA2_CLOBO STANDARD; PRT; 1295 AA.
ID BXA2_CLOBO
AC Q45894; P77780;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (Bont/A)
DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-

```





RX MEDLINE-93063293: PubMed-1331807;  
 RA Schiavo C., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RA Dasgupta B.R., Montecucco C.;  
 RL "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 by proteolytic cleavage of synaptobrevin";  
 Nature 359:832-835(1992).  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-1-PHE-77 BOND OF  
 SYNAPTOBREVIN-2.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M8186; AAA3231.1;  
 DR EMBL: Z11934; CA67399.1;  
 DR EMBL: X70817; CA560148.1;  
 DR EMBL: S07128; S07128.  
 DR PIR: S07125; S07125.  
 DR PIR: S08352; S08352.  
 DR PIR: S08573; S08573.  
 DR PIR: S08574; S08574.  
 DR PIR: A48940; A48940.  
 DR HSSP: P10845; 3BFA.  
 DR MEROPS: M27.002; -.  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR PRODOM: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 KM Neurotoxin: Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT INIT MET 0 0  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.  
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 436 445 INTERCHAIN (PROBABLE).  
 FT CONFLICT 29 29 T->M (IN REF. 4).  
 FT CONFLICT 217 217 R->G (IN REF. 2).  
 FT CONFLICT 224 224 A->S (IN REF. 2).  
 FT CONFLICT 463 463 S->R (IN REF. 4).  
 SQ SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;

Query Match 5.6%; Score 8; DB 1; Length 1290;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 NNSGWRIS 124  
 Db 957 NNSGWRIS 964

RESULT 6  
 ID BXA1\_CLOBO STANDARD; PRT; 1295 AA.  
 AC P10845; P18639; P01561.  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type A precursor (BC 3.4.24.69) (BONT/A)  
 DE (Bontoxilysin A) (BOTOX) [contains: Botulinum neurotoxin A, light-  
 DE chain; Botulinum neurotoxin A, heavy-chain].  
 GN BOTA OR BNA OR ATX.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 NCBI\_TaxID=1491;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=NCCTC 2916;  
 RX MEDLINE-90235864; PubMed-2185020;  
 RA Thompson D.E., Brehm J.K., Outram J.D., Swinfield T.-J.,  
 RA Shone C.C., Atkinson T., Melling J., Minton N.P.;  
 RA "The complete amino acid sequence of the Clostridium botulinum type A  
 RA neurotoxin, deduced by nucleotide sequence analysis of the encoding  
 RA gene";  
 RL Biochem. 189:73-81(1990).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=62A;  
 RX MEDLINE-90264400; PubMed-2160960;  
 RA Blitz B., Kiarzono H., Wille M., Frevent J., Wernars K., Niemann H.;  
 RA "The complete sequence of botulinum neurotoxin type A and comparison  
 RA with other clostridial neurotoxins";  
 RL J. Biol. Chem. 265:9153-9158(1990).  
 [3]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RP STRAIN=62A;  
 RX MEDLINE-97016817; PubMed-8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RA "Organization and phylogenetic interrelationships of genes encoding  
 RA components of the botulinum toxin complex in proteolytic Clostridium  
 RA botulinum types A, B, and F: evidence of chimeric sequences in the  
 RA gene encoding the nontoxic nonhemagglutinin component";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RP STRAIN=HALL;  
 RX MEDLINE-89350959; PubMed-2669749;  
 RA Boley M.J., Somers E., Dasgupta B.R.;  
 RA "Characterization of botulinum type A neurotoxin gene: delineation of  
 RA the N-terminal encoding region";  
 RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).  
 [5]  
 RP SEQUENCE OF 1-18 FROM N.A.  
 RP STRAIN=TYPE A NIH;  
 RX MEDLINE-9606783; PubMed-8521962;  
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
 RA "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
 RA components of Clostridium botulinum type A progenitor toxins";  
 RL FEBS Lett. 376:41-44(1995).  
 [6]  
 RP SEQUENCE OF 1-16.  
 RX MEDLINE-84178501; PubMed-6370252;  
 RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;  
 RA "Partial amino acid sequence of the heavy and light chains of  
 RA botulinum neurotoxin type A";  
 RL Biochem. Biophys. Res. Commun. 119:900-904(1984).  
 [7]  
 RP SEQUENCE OF 1-46.  
 RA Dasgupta B.R., Foley J., Niece R.;  
 RA "Partial sequence of the light chain of botulinum neurotoxin type A";  
 RL Biochemistry 26:4162-4162(1987).  
 [8]

RP SEQUENCE OF 1-5 AND 444-456.  
 RA MEDLINE-91120847; PubMed-2126206;  
 RX Dasgupta B.R., Dekleva M.L.;  
 RT "Botulinum neurotoxin type A: sequence of amino acids at the  
 RT N-terminus and around the nicking site.";  
 RL Biochimie 72:661-664(1990).  
 RN [9]  
 RP SEQUENCE OF 448-464 AND 872-895.  
 RX MEDLINE-89024662; PubMed-3178218;  
 RA Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;  
 RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two  
 RT halves and their partial sequences";  
 RL Arch. Biochem. Biophys. 266:142-151(1988).  
 RN [10]  
 RP SEQUENCE OF 448-482.  
 RX MEDLINE-85285016; PubMed-3896784;  
 RA Shone C.C., Hambleton P., Melling J.;  
 RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin  
 RT and purification of two tryptic fragments. Proteolytic action near  
 RT the COOH-terminus of the heavy subunit destroys toxin-binding  
 RT activity.";  
 RL Eur. J. Biochem. 151:75-82(1985).  
 RN [11]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE-94063091; PubMed-8243676;  
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
 RA Benfenati F., Wilson M.C., Montecucco C.;  
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 RT COOH-terminal peptide bonds";  
 RL FEBS Lett. 335:99-103(1993).  
 RN [12]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE-94124495; PubMed-8294407;  
 RA Binz T., Blaszi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 RA Jahn R., Niemann H.;  
 RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";  
 RL J. Biol. Chem. 269:1617-1620(1994).  
 RN [13]  
 RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.  
 RX PubMed-11700044;  
 RA Rigdon M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;  
 RT "Site-directed mutagenesis identifies active-site residues of the  
 RT light chain of botulinum neurotoxin type a.";  
 RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).  
 RN [14]  
 RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).  
 RX MEDLINE-98455071; PubMed-9783750;  
 RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;  
 RT "Crystal structure of botulinum neurotoxin type A and implications  
 RT for toxicity.";  
 RL Nat. Struct. Biol. 5:898-902(1998).  
 CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin  
 CC binds with high affinity to peripheral neuronal presynaptic  
 CC membrane, is then internalized by receptor-mediated endocytosis.  
 CC The C-terminus of the heavy chain (H) is responsible for the  
 CC adherence of the toxin to the cell surface while the N-terminus  
 CC mediates transport of the light chain from the endocytic vesicle  
 CC to the cytosol. After translocation, the light chain (L)  
 CC hydrolyzes the 197-Gln-1-Air-198 bond in SNAP-25, thereby blocking  
 CC neurotransmitter release. Inhibition of acetylcholine release  
 CC results in flaccid paralysis, with frequent heart or respiratory  
 CC failure.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a  
 CC heavy chain (H).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- PHARMACOLOGICAL: Available under the name BOTOX(R) (Allergan) for  
 CC the treatment of strabismus and blepharospasm associated with  
 CC dystonia and cervical dystonia. Also used for the treatment of  
 CC hemifacial spasm and a number of other neurological disorders  
 CC characterized by abnormal muscle contraction.

CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of  
 CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC EMBL; X52066; CA36289.1; -;  
 CC EMBL; M30196; AAA23262.1; -;  
 CC EMBL; X92973; CA63551.1; -;  
 CC EMBL; D67030; BAA11051.1; -;  
 CC EMBL; M27892; AAA23269.1; -;  
 CC PIR; A35294; B7CLAB.  
 CC PIR; S09492; S09492.  
 CC PDB; 3BT4; 01-OCT-99.  
 CC MEROPS; M27.002; -;  
 CC InterPro; IPR000395; Bontoxilysin.  
 CC InterPro; IPR000130; Zn\_MTPeptide.  
 CC Pfam; PF01742; Peptidase\_M27.1.  
 CC PRINTS; PR00760; BONTOTOXILYSIN.  
 CC PRODOM; PD001963; Bontoxilysin; 1.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;  
 CC Pharmaceutical; 3D-structure.  
 CC KW INT. MET 0  
 CC CHAIN 1 447  
 CC CHAIN 448 1295  
 CC METAL 222 222  
 CC ACT\_SITE 223 223  
 CC METAL 226 226  
 CC METAL 261 261  
 CC DISULFID 429 453  
 CC DISULFID 1234 1279  
 CC TRANSFID 626 646  
 CC TRANSMEM 655 675  
 CC VARIANT 26 26  
 CC MUTAGEN 261 261  
 CC FT MUTAGEN 265 265  
 CC FT MUTAGEN 365 365  
 CC FT CONFLICT 1 1  
 CC FT CONFLICT 479 479  
 CC FT CONFLICT 875 875  
 CC FT CONFLICT 891 891  
 CC SQ SEQUENCE 1295 AA; 149322 MW; 858342F754862579 CRC64;  
 CC  
 CC Query Match 5.6%; Score 8; DB 1; Length 1295;  
 CC Best Local Similarity 100.0%; Pred. No. 1.9;  
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 104 LNEYTII 111  
 CC Db 957 LNEYTII 964  
 CC  
 CC RESULT 7  
 CC MURF\_RICPR STANDARD; PRT; 449 AA.  
 CC ID MURF\_RICPR  
 CC AC 005953;  
 CC DT 30-MAY-2000 (Rel. 39, Created)  
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Probable UDP-N-acetylmuramoylalanyl-D-glutanyl-2,6-diaminopimelate--D-  
 CC alanyl-D-alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide  
 CC synthetase) (D-alanyl-D-alanine-adding enzyme).  
 CC GN MURF OR RP596.  
 CC OS Rickettsia prowazekii.

CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiaceae; Rickettsia.  
 CC NCBI\_TaxID=782;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E.  
 RX MEDLINE-99039499; PubMed-9823893;  
 RA Andersson S.G.E.; Zomrodipour A.; Andersson J.O.;  
 RA Scherzberg T.; Almaraz U.C.M.; Podowski R.M.; Naeslund A.K.;  
 RA Eriksson A.-S.; Winkler H.H.; Kurland C.G.;  
 RT The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.  
 RL Nature 396:133-140(1998).  
 RN (2)  
 RP SEQUENCE OF 1-96 FROM N.A.  
 RC STRAIN-MADRID E.  
 RX MEDLINE-97419517; PubMed-9274032;  
 RA Andersson J.O.; Andersson S.G.E.;  
 RT Genomic rearrangements during evolution of the obligate  
 RT intracellular parasite Rickettsia prowazekii as inferred from an  
 RT analysis of 52015 bp nucleotide sequence.  
 RL Microbiology 143:2783-2795(1997).  
 CC -1- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL  
 CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE  
 CC PRECURSOR OF MOREIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-  
 CC glutamate-meso-2,6-diaminopentanoate + D-alanyl-D-alanine = ADP  
 CC + carboxy-L-lysyl-D-N-acetylmuramoyl-L-alanyl-D-glutamy-6-  
 CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE MOREIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ235272; CA015040.1; -  
 DR EMBL; Y11783; CA072472.1; -  
 DR HSSP; P11880; ICG4.  
 DR InterPro; IPR000713; Mur\_Ligase.  
 DR InterPro; IPR004101; Mur\_Ligase\_C.  
 DR Pfam; PF02875; Mur\_Ligase\_1.  
 DR K1M; Peptidoglycan synthetase; Cell wall; Cell division; Ligase;  
 KM ATP-binding; Complete proteome.  
 FT NP\_BIND 106 112 ATP (POTENTIAL).  
 FT SEQUENCE 449 AA; 50672 MW; 3FEB8468F825BFD4 CRC64;  
 SO

Query Match 4.9%; Score 7; DB 1; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 NKFETDI 37  
 DB 268 NKFETDI 274

RESULT 8  
 VGLC\_HSYMB STANDARD; PRT; 501 AA.  
 ID VGLC\_HSYMB  
 AC P22650;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)  
 DE (GA).  
 GN GA.  
 RP

OS Marek's disease herpesvirus (strain bc-1) (MDHV).  
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Alphaherpesvirinae; Varicelloviruses.  
 CC NCBI\_TaxID=10387;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-90142542; PubMed-2559540;  
 RA Ihara T.; Kato A.; Ueda S.; Ishihama A.; Hirai K.;  
 RT Comparison of the sequence of the secretory glycoprotein A (ga) gene  
 RT in MDs and BC-1 strains of Marek's disease virus type 1.  
 RL Virus Genes 3:127-140(1989).  
 CC -1- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF  
 CC MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE  
 CC IMMUNOSUPPRESSION THAT OCCURS AFTER MDV INFECTION.  
 CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT  
 CC OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY  
 CC OTHER INTERACTIONS.  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUS GLYCOPROTEIN C FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D90002; BA014054.1; ALT\_SEQ.  
 DR PIR; J50388; VGPBPA.  
 DR InterPro; IPR001654; Marek\_A.  
 DR Pfam; PF02124; Marek\_A\_1.  
 DR PRINTS; PR00675; MAREKSGPA.  
 DR Glycoprotein; Transmembrane; Signal.  
 KW SIGNAL 1 27  
 FT CHAIN 28 501  
 FT TRANSMEM 466 492  
 FT CARBOHYD 46 46  
 FT CARBOHYD 91 100  
 FT CARBOHYD 100 101  
 FT CARBOHYD 120 120  
 FT CARBOHYD 120 122  
 FT CARBOHYD 212 212  
 FT CARBOHYD 354 354  
 FT CARBOHYD 400 400  
 FT CARBOHYD 429 429  
 FT CARBOHYD 493 493  
 FT SEQUENCE 501 AA; 56134 MW; D671E952331024B0 CRC64;  
 SO

Query Match 4.9%; Score 7; DB 1; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 SNISNG 48  
 DB 45 SNISNG 51

RESULT 9  
 VGLC\_HSYMD STANDARD; PRT; 501 AA.  
 ID VGLC\_HSYMD  
 AC P33500;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)  
 DE (GA).  
 GN GA.  
 OS Marek's disease herpesvirus (strain RB-1B) (MDHV).  
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Alphaherpesvirinae; Varicelloviruses.  
 CC NCBI\_TaxID=33707;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RX MEDLINE-89269090; PubMed=2543160;  
RA Blum M.M., Ross N.L.U.;  
RT "Nucleotide sequence of the Marek's disease virus (MDV) RB-1B A  
RT antigen gene and the identification of the MDV A antigen as the  
RT herpes simplex virus-1 glycoprotein C homologue";  
RL Virus Res. 12:371-382(1989).  
CC -1- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF  
CC MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE  
CC IMMUNOSUPPRESSION THAT OCCURS AFTER MDV INFECTION.  
CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT  
CC OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY  
CC OTHER INTERACTIONS.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.  
DR PIR: A60005; A60005.  
DR InterPro: IPR001654; Marek\_A.  
DR Pfam: PF02124; Marek\_A.1.  
DR PRINTS: PR00675; MAREKSPA.  
KM Glycoprotein; Transmembrane; Signal.  
FT SIGNAL 1 27  
FT CHAIN 1 27  
FT TRANSMEM SECRETORY GLYCOPROTEIN GP57-65.  
FT TRANSMEM 466 501  
FT CARBOHYD 46 492  
FT CARBOHYD 91 91  
FT CARBOHYD 100 100  
FT CARBOHYD 120 120  
FT CARBOHYD 120 120  
FT CARBOHYD 212 212  
FT CARBOHYD 354 354  
FT CARBOHYD 400 400  
FT CARBOHYD 429 429  
FT CARBOHYD 493 493  
SQ SEQUENCE 501 AA; 56104 MW; 438473BDD79340A CRC64;  
  
Query Match 4.9%; Score 7; DB 1; Length 501;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 42 SNISING 48  
Db 45 SNISING 51  
|||||  
ID VGLC\_HSVGM STANDARD; PRT; 501 AA.  
AC P22651;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)  
DE (GA).  
GN GA.  
OS Marek's disease herpesvirus (strain Md5) (MDHV).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=10389;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-90142542; PubMed=2559540;  
RA Ihara T., Kato A., Ueda S., Ishihama A., Hirai K.;  
RT "Comparison of the sequence of the secretory glycoprotein A (ga) gene  
RT in Md5 and BC-1 strains of Marek's disease virus type 1";  
RL Virus Genes 3:127-140(1989).  
CC -1- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF  
CC MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE  
CC IMMUNOSUPPRESSION THAT OCCURS AFTER MDV INFECTION.  
CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT  
CC OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY  
CC OTHER INTERACTIONS.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
CC -----  
DR EMBL: D90001; BAA14052.1; -.  
DR PIR: A22818; VGBMB.  
DR InterPro: IPR001654; Marek\_A.  
DR Pfam: PF02124; Marek\_A.1.  
DR PRINTS: PR00675; MAREKSPA.  
KM Glycoprotein; Transmembrane; Signal.  
FT SIGNAL 1 27  
FT CHAIN 1 27  
FT TRANSMEM SECRETORY GLYCOPROTEIN GP57-65.  
FT TRANSMEM 466 492  
FT CARBOHYD 46 46  
FT CARBOHYD 91 91  
FT CARBOHYD 100 100  
FT CARBOHYD 100 100  
FT CARBOHYD 120 120  
FT CARBOHYD 212 212  
FT CARBOHYD 354 354  
FT CARBOHYD 400 400  
FT CARBOHYD 429 429  
FT CARBOHYD 493 493  
SQ SEQUENCE 501 AA; 36088 MW; 4393C56AA779340A CRC64;

Query Match 4.9%; Score 7; DB 1; Length 501;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 42 SNISING 48  
Db 45 SNISING 51  
|||||  
ID VGLC\_HSVGM STANDARD; PRT; 505 AA.  
AC P10681;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)  
DE (GA).  
GN GA.  
OS Marek's disease herpesvirus (strain GA) (MDHV).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=10388;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-88230597; PubMed=2836620;  
RA Cousens P.M., Velicer L.F.;  
RT "Structure and complete nucleotide sequence of the Marek's disease  
RT herpesvirus gp57-65 gene";  
RL J. Virol. 62:2373-2379(1988).  
CC -1- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF  
CC MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE  
CC IMMUNOSUPPRESSION THAT OCCURS AFTER MDV INFECTION.  
CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT  
CC OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY  
CC OTHER INTERACTIONS.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
CC -----

```

DR EMBL: M20001; AAA46114.1; -.
DR PIR: A28843; VGBEMH.
DR InterPro: IPR001654; Marek_A.
DR Pfam: PF02124; Marek_A; 1.
DR PRINTS: PR00675; MAREKSGPA.
DR Glycoprotein; Transmembrane; Signal.
KW SIGNAL.
FT CHAIN 1 27 POTENTIAL.
FT TRANSMEM 28 505 SECRETORY GLYCOPROTEIN GP57-65.
FT CARBOHYD 465 491 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 505 AA; 56809 MW; D06D75D7D9C666D CRC64;

Query Match 4.9%; Score 7; DB 1; Length 505;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 SNISING 48
Db 44 SNISING 50

RESULT 12
RRBL_YEAST STANDARD; PRT; 511 AA.
ID RRBL_YEAST
AC 004225;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Ribosome assembly protein RRBL.
GN RRBL OR YMR131C OR YM9553.07C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Badcock K., Chutcher C., Bartell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RA CHARACTERIZATION.
RA MEDLINE=21585391; PubMed=11728313;
RA Schaper S., Fromont-Rachine M., Linder P., de la Cruz J., Naman A.,
RA Yaniv M.;
RT "A yeast homolog of chromatin assembly factor 1 is involved in early
RT ribosome assembly."
RL Curr. Biol. 11:1885-1890(2001).
CC -!- FUNCTION: Involved in regulation of L3 expression and stability
CC and plays a role in early 60S ribosomal subunit assembly. May be
CC required for proper assembly of preribosomal particles during
CC early ribosome biogenesis, presumably by targeting L3 onto the 35S
CC precursor rRNA.
CC -!- SUBUNIT: Associates with ribosomal protein L3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 4 WD REPEATS (TTP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z48622; CAA8556.1; -.

```

```

DR SGD: S0004738; RRBL.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 4.
DR PRINTS: PR00320; GPROTEINRPT.
DR SMART: SM00320; WD40; 4.
DR PROSITE: PS00678; WD_REPEATS_1; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Ribosome biogenesis; rRNA processing; Nuclear protein; Repeat;
KW WD repeat.
FT REPEAT 319 359 WD 1.
FT REPEAT 364 404 WD 2.
FT REPEAT 415 455 WD 3.
FT REPEAT 477 510 WD 4.
SQ SEQUENCE 511 AA; 57261 MW; 1D18CE3C60BAFF30 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 511;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 NGDYVYI 53
Db 235 NGDYVYI 241

RESULT 13
POLG_EMCVB STANDARD; PRT; 2292 AA.
ID POLG_EMCVB
AC P17593;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein (Contains: Coat proteins VP1 TO VP4; Core proteins
DE EC2 TO P2C; P2A: Genome-linked protein VP3; Picornain 3C
DE EC3 4.22.28) (Protease 3C) (P2C). RNA-directed RNA polymerase P2D
DE (EC 2.7.7.48)
OS Encephalomyocarditis virus (strain emc-b nondiabetogenic)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Cardioviruses
OX NCBI_TaxID=12105;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=69243189; PubMed=2541543;
RA Bae Y.S., Eun H.W., Yoon J.W.;
RT "Genomic differences between the diabetogenic and nondiabetogenic
RT variants of encephalomyocarditis virus."
RL Virology 170:282-287(1989).
CC -!- FUNCTION: P2C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M22457; AAA43033.1; ALT_SEQ.
DR PIR: B31473; GNNYEB.
DR HSSP: P12296; 2MEV.
DR MEROPS: C03.009; -.
DR MEROPS: U29.001; -.
DR InterPro: IPR001605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; RNV.
DR Pfam: PF00073; RNV; 3.

```

DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.  
FT PROPEP 1 67 LEADER PEPTIDE.  
FT CHAIN 68 137 COAT PROTEIN VP4 (RHQ).  
FT CHAIN 138 393 COAT PROTEIN VP2 (BETA).  
FT CHAIN 394 624 COAT PROTEIN VP3 (GAMMA).  
FT CHAIN 625 901 COAT PROTEIN VP1 (ALPHA).  
FT CHAIN 902 1058 CORE PROTEIN P2A (G).  
FT CHAIN 1059 1194 CORE PROTEIN P2B (I).  
FT CHAIN 1195 1519 CORE PROTEIN P2C (F).  
FT CHAIN 1520 1607 CORE PROTEIN P3A.  
FT CHAIN 1608 1627 GENOME-LINKED PROTEIN VPG (H).  
FT CHAIN 1628 1832 PICORNAIN 3C (P22).  
FT CHAIN 1833 2292 RNA-DIRECTED RNA POLYMERASE P3D (E).  
FT LIPID 68 68 MYRISTATE (BY SIMILARITY).  
FT ACT\_SITE 1786 1786 PROTEASE (POTENTIAL).  
FT ACT\_SITE 1804 1804 PROTEASE (POTENTIAL).  
SQ SEQUENCE 2292 AA; 255495 MW; 8340DDB1437B8D4 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 2292;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 134 LQDTAGN 140  
Db 152 LQDTAGN 158

RESULT 14  
POLG-EMCVD STANDARD; PRT; 2292 AA.  
ID POLG-EMCVD  
AC P17594;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; P3A; Genome-linked protein VPG; Picornain 3C  
DE (EC 3.4.22.28) (Protease 3C); RNA-directed RNA polymerase P3D  
DE (EC 2.7.7.48)].  
OS Encephalomyocarditis virus (strain emc-d diabetogenic).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Cardiovirus.  
OX NCBI\_TaxID=12106;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89243189; PubMed=2541543;  
RA Bae Y.S., Eun H.M., Yoon J.W.;  
RT "Genomic differences between the diabetogenic and nondiabetogenic  
RT variants of encephalomyocarditis virus.";  
RL Virology 170:282-287(1989).  
CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN  
CC O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M22458; AAA43034.1; -  
DR PIR; A31473; GNNYD.  
DR HSP; P12296; 2MEV.  
DR MEROPS; C03.009; -.

DR MEROPS; U29.001; -  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR001676; RHV.  
DR Pfam; PF00073; rhv; 3.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polypeptide; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.  
FT PROPEP 1 67 LEADER PEPTIDE.  
FT CHAIN 68 137 COAT PROTEIN VP4 (RHQ).  
FT CHAIN 138 393 COAT PROTEIN VP2 (BETA).  
FT CHAIN 394 624 COAT PROTEIN VP3 (GAMMA).  
FT CHAIN 625 901 COAT PROTEIN VP1 (ALPHA).  
FT CHAIN 902 1058 CORE PROTEIN P2A (G).  
FT CHAIN 1059 1194 CORE PROTEIN P2B (I).  
FT CHAIN 1195 1519 CORE PROTEIN P2C (F).  
FT CHAIN 1520 1607 CORE PROTEIN P3A.  
FT CHAIN 1608 1627 GENOME-LINKED PROTEIN VPG (H).  
FT CHAIN 1628 1832 PICORNAIN 3C (P22).  
FT CHAIN 1833 2292 RNA-DIRECTED RNA POLYMERASE P3D (E).  
FT LIPID 68 68 MYRISTATE (BY SIMILARITY).  
FT ACT\_SITE 1786 1786 PROTEASE (POTENTIAL).  
FT ACT\_SITE 1804 1804 PROTEASE (POTENTIAL).  
SQ SEQUENCE 2292 AA; 255426 MW; F2B0627B0F444107 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 2292;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 134 LQDTAGN 140  
Db 152 LQDTAGN 158

RESULT 15  
CATR-HUMAN STANDARD; PRT; 79 AA.  
ID CATR-HUMAN  
AC Q13166;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE CATR tumorigenic conversion 1 protein (CATR1.3).  
DE CATR1.  
GN CATR1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cardioma;  
RX MEDLINE=95327656; PubMed=7604004;  
RA Li D., Noyes I., Shuler C., Mito G.E.;  
RT "Cloning and sequencing of CATR1.3, a human gene associated with  
RT tumorigenic conversion.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:6409-6413(1995).  
CC -1- DEVELOPMENTAL STAGE: ASSOCIATED WITH TUMORIGENIC CONVERSION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U25433; -; NOT\_ANNOTATED\_CDS.  
DR MIM; 600676; -  
SQ SEQUENCE 79 AA; 9224 MW; BC3667C05911ACF3 CRC64;

Query Match 4.2%; Score 6; DB 1; Length 79;

Best Local Similarity 100.0%; Pred. NO. 24;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 QNNDII 79  
Db 15 QNNDII 20

Search completed: August 15, 2002, 11:24:38  
Job time: 685 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:26 ; Search time 96.53 Seconds  
(without alignments)  
165,696 Million cell updates/sec

Title: US-08-981-087a-3  
Perfect score: 144  
Sequence: 1 VENTYOMISISDYINKMFV.....ITONSFLNINOGRGYOKP 144

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 11073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq.032802.\*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	100.0	144	18	AAW09016
2	144	100.0	431	18	AAW09014
3	144	100.0	432	21	AAW77138
4	144	100.0	432	22	AAW04096
5	144	100.0	432	22	AAW04103
6	144	100.0	645	22	AAW07894
7	144	100.0	685	22	AAW07893
8	144	100.0	862	22	AAW07890
9	144	100.0	887	22	AAW07892
10	144	100.0	1032	22	AAW07901
11	144	100.0	1059	21	AAW93309

12	144	100.0	1084	21	AAW93312	A. manganease supero	
13	144	100.0	1092	22	AAE07800	C. botulinum C2 tr	
14	69	47.9	660	22	AAE07898	Modified clostridi	
15	27	18.8	448	19	AAW68399	Clostridium botulli	
16	15	10.4	419	22	AAW04095	Botulinm toxin hea	
17	15	10.4	449	21	AAW77137	Synthetic botulinu	
18	15	10.4	449	22	AAW04094	Botulinum toxin hea	
19	15	10.4	451	19	AAW68395	Clostridium botulli	
20	15	10.4	452	19	AAW68396	Clostridium botulli	
21	21	11	233	21	AAW77143	Botulinum neurotox	
22	22	11	352	21	AAW36303	BONT/A prototoxin	
23	23	11	7.6	415	22	AAW04083	Botulinm toxin C f
24	24	11	7.6	432	21	AAW77142	Native botulinum n
25	25	11	7.6	434	22	AAW04089	Botulinm toxin hea
26	26	11	7.6	435	22	AAW04090	Botulinm toxin hea
27	27	11	7.6	437	22	AAW04088	Botulinm toxin hea
28	28	11	7.6	438	17	AAW95008	Type A neurotoxin
29	29	11	7.6	438	19	AAW68389	Clostridium botulli
30	30	11	7.6	438	21	AAW77134	Synthetic botulinu
31	31	11	7.6	445	19	AAW68391	Clostridium botulli
32	32	11	7.6	462	17	AAW95009	Type A neurotoxin
33	33	11	7.6	462	19	AAW68390	Clostridium botulli
34	34	11	7.6	837	21	AAW77140	Native botulinum n
35	35	11	7.6	847	22	AAW04081	Botulinm toxin hea
36	36	11	7.6	1067	21	AAW93307	A. manganease supero
37	37	11	7.6	1092	21	AAW93310	C. botulinum type
38	38	11	7.6	1296	17	AAW93010	Synthetic botulinu
39	39	8	5.6	449	21	AAW77139	Botulinm toxin hea
40	40	8	5.6	449	22	AAW04167	Clostridium botulli
41	41	8	5.6	473	19	AAW68400	Botulinm toxin hea
42	42	7	4.9	20	16	AAW64982	MLV p15E C19 bind
43	43	7	4.9	196	15	AAW43986	p15E protein. Hum
44	44	7	4.9	196	15	AAW53443	Sequence of envelo
45	7	4.9	206	21	AAW77144	Botulinum neurotox	

ALIGNMENTS

RESULT 1  
AAW09016  
ID AAW09016 standard; Protein: 144 AA.  
XX  
AC AAW09016;  
XX  
XX 31-MAR-1997 (first entry)  
XX  
XX Immunogenic type F botulinum toxin polypeptide (aa992-1135).  
XX  
XX Botulinum toxin; neurotoxin; BONT/F; immunogen; vaccine; botulism.  
XX  
XX Clostridium botulinum type F strain Langeland.  
XX  
XX  
XX W09641881-A1.  
XX  
XX 27-DEC-1996.  
XX  
XX 12-JUN-1996; 96WO-GB01409.  
XX  
XX 12-JUN-1995; 95GB-0011909.  
XX  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
XX Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;  
XX WPI; 1997-065467/06.  
XX  
XX Immunogenic type F botulinum toxin polypeptide(s) - allows  
XX PT recombinant vaccine produ.  
XX  
XX Claim 5; Page 18-19; 37pp; English.  
XX  
XX Novel polypeptides (AAW09014-17) respectively comprise amino acids

CC 848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a  
 CC type F botulinum neurotoxin (BoNT/F). They lack the L chain and  
 CC HN epitopes necessary for metalloprotease activity and toxin  
 CC internalisation. They are free of botulinum toxin activity but can  
 CC induce protective immunity to a type F botulinum toxin, making them  
 CC useful for vaccine prodn. Recombinant polypeptides can be  
 CC produced in transformed host cells, esp. as fusion proteins, e.g.  
 CC with maltose binding protein to facilitate purification.  
 SQ Sequence 144 AA;

Query Match 100.0%; Score 144; DB 18; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-147;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VENTOMISIDYINKWFVTITNNRLGNSRIYINGNLIDKESISNIGDIHVSNDITFKI 60  
 Db 1 vfnycgmisdsdylnkwlfvcltmrfgnsrilyngnlidksisnlgdihvsdnllfki 60

QY 61 VGCNDTRVVGIRYRKVPEDTELKTEIETLYSDEDDPSILKDFWGNLYLLNRRYLLNLLR 120  
 Db 61 vgcndtrvygiryrfkvtelgkteielysdepdpssllkdfwgnlyllnkrlyllnllr 120

QY 121 TDKSITQNSNLFNLINQGRGYQKP 144  
 Db 121 tdksitqnsnflnlnqgrgyqkp 144

RESULT 2  
 ID AAM09014 standard; Protein; 431 AA.  
 XX AAM09014;  
 AC AAM09014;  
 DT 31-MAR-1997 (first entry)  
 DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).  
 XX  
 KW Botulinum toxin; neurotoxin; BoNT/F; Immunogen; vaccine; botulism.  
 KM  
 OS Clostridium botulinum type F strain Langeland.  
 XX  
 PN W09641881-A1.  
 PD 27-DEC-1996.  
 XX  
 PF 12-JUN-1996; 96WO-GB01409.  
 XX  
 PR 12-JUN-1995; 95GB-0011909.  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;  
 DR WPI: 1997-065467/06.  
 DR N-PSDB: AAT48100.  
 PT Immunogenic type F botulinum toxin polypeptide(s) - allows  
 PT recombinant vaccine prodn.  
 XX  
 PS Claim 5; Page 16-17; 37pp; English.  
 XX  
 CC A polypeptide (AAM09014) comprises the heavy chain (amino acids  
 CC 848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be  
 CC produced using a synthetic gene (AAT48101) based on the natural  
 CC gene sequence (AAT48100) for the heavy chain. The polypeptides and  
 CC its fragments (see also AAM09015-17) lack the light chain and HN  
 CC epitopes necessary for metalloprotease activity and toxin  
 CC internalisation. They are free of botulinum toxin activity but can  
 CC induce protective immunity to a type F botulinum toxin, making them  
 CC useful for vaccine prodn. Recombinant polypeptides can be  
 CC produced in transformed host cells, esp. as fusion proteins, e.g.

CC with maltose binding protein to facilitate purification.  
 SQ Sequence 431 AA;

Query Match 100.0%; Score 144; DB 18; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-147;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VENTOMISIDYINKWFVTITNNRLGNSRIYINGNLIDKESISNIGDIHVSNDITFKI 60  
 Db 145 vfnycgmisdsdylnkwlfvcltmrfgnsrilyngnlidksisnlgdihvsdnllfki 204

QY 61 VGCNDTRVVGIRYRKVPEDTELKTEIETLYSDEDDPSILKDFWGNLYLLNRRYLLNLLR 120  
 Db 205 vgcndtrvygiryrfkvtelgkteielysdepdpssllkdfwgnlyllnkrlyllnllr 264

QY 121 TDKSITQNSNLFNLINQGRGYQKP 144  
 Db 265 tdksitqnsnflnlnqgrgyqkp 288

RESULT 3  
 ID AAT77138 standard; Protein; 432 AA.  
 XX AAT77138;  
 AC AAT77138;  
 DT 08-MAY-2000 (first entry)  
 DE Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.  
 XX  
 KW Botulinum neurotoxin; heavy chain; BoNT; serotype F;  
 KM C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
 KW VEE; botulism; vaccine; diagnosis; drug screening.  
 XX  
 OS Clostridium botulinum.  
 OS Synthetic.  
 PN W0200002524-A2.  
 PD 20-JAN-2000.  
 XX  
 PF 09-JUL-1999; 99WO-US15570.  
 XX  
 PR 10-JUL-1998; 98US-0092416.  
 PR 12-MAY-1999; 99US-0133870.  
 XX  
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith LJ;  
 DR WPI: 2000-160827/14.  
 DR N-PSDB: AA287216.  
 PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 PT toxin serotypes A-G, is used for inducing an immune response against  
 PT botulinum -  
 XX  
 PS Claim 27; Page -; 54pp; English.  
 XX  
 CC The invention relates to novel vaccines that induce a protective immune  
 CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
 CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
 CC DNA construct comprising a vector, and at least one nucleic acid  
 CC fragment comprising a C-terminal heavy chain fragment (HC) from BoNT  
 CC serotypes A-G. In preferred embodiments of the invention, the vector is  
 CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
 CC this vector results in the production of large amounts of a protein  
 CC encoded by a sequence cloned into the replicon. The constructs are used  
 CC to produce vaccines against botulism. The proteins can also be used as  
 CC diagnostic tools for the diagnosis of botulism. The transformed host  
 CC cells can be used to analyse the effectiveness of drugs and agents which

inhibitory effects. The vaccine currently used against botulism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine of overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AAV7134-Y71139 represent synthetic Bont Hc fragments used in the present invention. The DNA encoding these sequences had been optimised for codon usage for expression in yeast. Note: This sequence is not given in the specification, but is decoded from the Bont Hc DNA sequence given on pages 45-46.

Sequence 432 AA;

Query Match	100.0%	Score 144;	DB 21;	Length 432;
Best Local Similarity	100.0%	Pred. No. 3.9e-147;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

**OY**    1 VENTQIMISIDYINKWIEVTTNNRLCSRIYINGNLIDEKSISNGLDIHVSNDIIFKI 60  
**Dd**    146 vfnytqmisisdyinkwifvltlnrnlsrlyngnlidexsisnlgdlhvsnllfki 209

**QY** 61 VGCNDTRVVGIRYRKVEDTELCKTEIEIETLYSDPEDPSILKDFWGNLLYNKRYYLLNLR 120  
**Db** 206 vgcndtrvgiryrkvdteiqkteielysdcdpsilkdftwgnllynkrzyllnlr 265

```

Oy      121 TDKSITONSFLINQOQGVYQKP 144
         |||||
Db      266 tdkstqnsflnlnqgrgyqkp 285

```

RESULT	4
AAB04096	
ID	AAB04096 standard; Protein; 432 AA.
YV	

AC	AAB04096;
XX	
DT	11-APR-2001 (first entry)
XX	

DE Botulism toxin heavy chain C-terminal sequence (serotype F).  
XX

KM Botulism; toxin; neurotoxin; heavy chain; recombinant expression  
 KM recombinant vector; antigen; immune response; vaccine; bacterium;  
 KM infection.

OS Synthetic.  
OS Clostridium botulinum.

PN WO200067700-A2.

PD 16-NOV-2000.

PF 12-MAY-2000; 2000WO-US12890.

PR 12-MAY-1999; 99US-0133865  
PR 12-MAY-1999; 99US-0133867  
PR 12-MAY-1999; 99US-0133867  
PR 12-MAY-1999; 99US-0133868  
PR 12-MAY-1999; 99US-0133868  
PR 12-MAY-1999; 99US-0133873  
PR 12-MAY-1999; 99US-0133873  
PR 29-JUL-1999; 99US-0146192

PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.

PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H, ...

DR WPI; 2001-016048/02.

XX

**PT** New nucleic acids encoding the carboxy- or amino-terminal portions of the heavy chain of botulinum neurotoxin of serotype A-G, useful as

PT	vaccine against botulism
XX	
PS	Claim 3; Fig 9b; 73pp; English
XX	

CC Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked forming a dimer chain  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (NH) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organisms such as *Escherichia coli* or *Pichia pastoris*. The use  
CC of the recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered producing  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from *Clostridium botulinum* serotypes A-G.

Sequence 432 AA:

Query Match	100.0%;	Score 144;	DB 22;	length 432;
Best Local Similarity	100.0%;	Pred. No. 3.9e-147;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

OY 61 VGCNDTRYVGRIRFKAFDTELCKTEIEFLYSDEDPSSILKDFMGNYLLNKKRYLLNLRL 1200  
|||||  
Db 206 vgcndtryvgirfkafdtelqkteietlgsdepssilkdmgnyllnkkryllnlrl 2655

```
Qy 121 TDKSITQNSNFIINQQRGVYQKP 144
    |||||
Db 266 tdkstqnsfnlningrgvvyqkp 289
```

RESULT	5
AAB04103	
ID	AAB04103 standard; Protein; 432 AA

AC AAB04103

DT 11-APR-2001 (first entry)

DE Botulism toxin heavy chain C-terminal sequence (serotype F)

Botulism; toxin; neurotoxin; heavy chain; recombinant expression;  
KW recombinant vector; antigen; immune response; vaccine; bacterium;  
KW infection.

OS Synthetic.  
OS Clostridium botulinum

PN WO200067700-A2

16-NOV-2000 PD

XX	12-MAY-1999;	99US-0133865
PR	12-MAY-1999;	99US-0133866
PR	12-MAY-1999;	99US-0133867
PR	12-MAY-1999;	99US-0133868
PR	12-MAY-1999;	99US-0133869
PR	12-MAY-2000;	2000WO-US12890

PA	(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX	
FN	23 00Z 1999, 2003 014010Z.

XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
XX WPI: 2001-016048/02.  
DR N-PSDB; AAA54499.  
XX  
XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
PT vaccine against botulism  
XX  
XX Disclosure; Fig 18b; 73pp; English.  
XX  
XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked, forming a dichain  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.  
XX  
SQ Sequence 432 AA;

Query Match 100.0%; Score 144; DB 22; Length 432;  
Best Local Similarity 100.0%; Pred. No. 3.9e-147;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VENTYOMISISDYINKMEFTITNNRLGNSRIYINGNLIDEKTSISNGDIHVSNDILFKI 60  
Db 146 VFNYTQMISISDYINKWLFVLTNNRIGNSRIYINGNLIDEKTSISNGDIHVSNDILFKI 205  
QY 61 VGCNDTRYVGIIRYKRVDTTELKTEIEFLYSDEPDPSILKDFMGNLYLLNKRYYLLNLR 120  
Db 206 VGCNDTRYVGIIRYKRVDTTELKTEIEFLYSDEPDPSILKDFMGNLYLLNKRYYLLNLR 265  
QY 121 TDKSITONSNFLINQOGRGYQKP 144  
Db 266 TDKSITONSNFLINQOGRGYQKP 289

## RESULT 6

AAE07894  
ID AAE07894 standard; Protein; 645 AA.  
XX  
AC AAE07894;  
XX  
DT 01-NOV-2001 (first entry)  
XX  
DE Modified clostridial heavy chain fragment #1.  
XX  
KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
XX diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.  
OS Chimeric - Corynebacterium diphtheriae.  
XX  
OS Chimeric - Clostridium botulinum.  
XX  
PN WO200158936-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 04-DEC-2000; 2000WO-GB04644.  
XX  
PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.  
XX  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
XX Shone CC, Sutton JM, Silman N;  
XX  
XX WPI: 2001-514643/56.  
DR  
XX  
XX New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -  
XX  
XX Example 2; Page 44; 50pp; English.  
XX  
XX The invention relates to a non toxic polypeptide, for delivery of a  
CC therapeutic agent to a neuronal cell, which comprises a binding domain  
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
CC as HC) that binds to the neuronal cell and a translocation domain (amino  
CC terminal half of HC, designated as HN), that translocates the therapeutic  
CC agent into the neuronal cell, where the translocation domain is not a HN  
CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
CC useful for the treatment of a disease state associated with neuronal  
CC cells. The polypeptide constructs are useful for delivering therapeutic  
CC substances to neuronal cells. They are useful to treat disorders of the  
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
CC and infection. They are also useful in gene therapy. The present sequence  
CC is modified clostridial heavy chain fragment. This sequence is  
CC constructed by fusing the binding domain of botulinum neurotoxin type F  
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.  
XX  
SQ Sequence 645 AA;

Query Match 100.0%; Score 144; DB 22; Length 645;  
Best Local Similarity 100.0%; Pred. No. 5.7e-147;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VENTYOMISISDYINKMEFTITNNRLGNSRIYINGNLIDEKTSISNGDIHVSNDILFKI 60  
Db 359 VFNYTQMISISDYINKWLFVLTNNRIGNSRIYINGNLIDEKTSISNGDIHVSNDILFKI 418  
QY 61 VGCNDTRYVGIIRYKRVDTTELKTEIEFLYSDEPDPSILKDFMGNLYLLNKRYYLLNLR 120  
Db 419 VGCNDTRYVGIIRYKRVDTTELKTEIEFLYSDEPDPSILKDFMGNLYLLNKRYYLLNLR 478  
QY 121 TDKSITONSNFLINQOGRGYQKP 144  
Db 479 TDKSITONSNFLINQOGRGYQKP 502

## RESULT 7

AAE07893  
ID AAE07893 standard; Protein; 685 AA.  
XX  
AC AAE07893;  
XX  
DT 01-NOV-2001 (first entry)  
XX  
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.  
XX  
XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
XX superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.  
XX  
OS Chimeric - Bacillus stearothermophilus.  
OS Chimeric - Influenza virus.  
OS Chimeric - Clostridium botulinum.  
XX  
OS Chimeric - Synthetic.  
XX  
PN WO200158936-A2.  
XX  
PD 16-AUG-2001.

```

XX 04-DEC-2000; 2000MO-GB04644.
PF 02-DEC-1999; 99GB-0028530.
XX 07-APR-2000; 2000GB-0008658.
PR (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX PI Shone CC, Sutton JM, Silman N;
XX WPI: 2001-514643/56.
XX DR
XX PT New non toxic polypeptide for delivery of a therapeutic agent for the
XX PT treatment of a CNS disorder comprising a binding domain that
XX PT translocates the therapeutic agent into the neuronal cells -
XX PS Example 9; Page 43; 50pp; English.
XX CC The invention relates to a non toxic polypeptide, for delivery of a
XX CC therapeutic agent to a neuronal cell, which comprises a binding domain
XX CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
XX CC as HC) that binds to the neuronal cell and a translocation domain (amino
XX CC terminal half of HC, designated as HN), that translocates the therapeutic
XX CC agent into the neuronal cell, where the translocation domain is not a HN
XX CC domain of a clostridial neurotoxin and is not a fragment or derivative of
XX CC a HN domain of a clostridial toxin. Polypeptides of the invention are
XX CC useful for the treatment of a disease state associated with neuronal
XX CC cells. The polypeptide constructs are useful for delivering therapeutic
XX CC substances to neuronal cells. They are useful for treating disorders of the
XX CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
XX CC and infection. They are also useful in gene therapy. The present sequence
XX CC is modified clostridial heavy chain-superoxide dismutase conjugate. This
XX CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
XX CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
XX CC translocation peptide from Influenza Virus and a neuronal cell-specific
XX CC binding domain from Botulinum neurotoxin type F (BoNT/F).
XX SQ Sequence 685 AA:
XX
XX Query Match 100.0%; Score 144; DB 22; Length 685;
XX Best Local Similarity 100.0%; Pred. No. 6e-147;
XX Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 VENVYOMISIDSYINKKAFVITNNRNLGNSRIYINGNLIDKSSISNLDIHSNDILFKI 60
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 399 vfnvgmstisdyinkwlfvltltnrlngrslyingnlidkssisnldhvsdnllfki 458
XX
XX QY 61 VGCNDTRYVGIRFKFVDELTGKTEIETLYSDPPDSILKDFMGNYLLYKRRYLLNLRL 120
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 459 vgcndtryvgirfkfvtelgkteleclysdepdpsllkdtwgnyllynkrryllnllr 518
XX
XX QY 121 TDKSTQNSNPLINQGRGYOKP 144
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 519 tdkstqnsnplinqgrgyvqkp 542
XX
XX RESULT 8
XX AAEO7890 standard; Protein: 862 AA.
XX ID AAEO7890;
XX AC AAEO7890;
XX DT 01-NOV-2001 (first entry)
XX DE Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
XX XX tumor; infection; neurodegenerative disease; gene therapy; chimeric;
XX XX superoxide dismutase; SOD; diphtheria neurotoxin;
XX XX botulinum neurotoxin type F; BoNT/F.
XX OS Chimeric - Bacillus stearothermophilus.

```

```

OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
XX OS Chimeric - Synthetic.
XX FM W0200158936-A2.
XX PD 16-AUG-2001.
XX PF 04-DEC-2000; 2000MO-GB04644.
XX 02-DEC-1999; 99GB-0028530.
XX 07-APR-2000; 2000GB-0008658.
XX PR (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX PA Shone CC, Sutton JM, Silman N;
XX PI WPI: 2001-514643/56.
XX DR
XX XX New non toxic polypeptide for delivery of a therapeutic agent for the
XX PT treatment of a CNS disorder comprising a binding domain that
XX PT translocates the therapeutic agent into the neuronal cells -
XX PS Example 9; Page 40; 50pp; English.
XX CC The invention relates to a non toxic polypeptide, for delivery of a
XX CC therapeutic agent to a neuronal cell, which comprises a binding domain
XX CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
XX CC as HC) that binds to the neuronal cell and a translocation domain (amino
XX CC terminal half of HC, designated as HN), that translocates the therapeutic
XX CC agent into the neuronal cell, where the translocation domain is not a HN
XX CC domain of a clostridial neurotoxin and is not a fragment or derivative of
XX CC a HN domain of a clostridial toxin. Polypeptides of the invention are
XX CC useful for the treatment of a disease state associated with neuronal
XX CC cells. The polypeptide constructs are useful for delivering therapeutic
XX CC substances to neuronal cells. They are useful for treating disorders of the
XX CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
XX CC and infection. They are also useful in gene therapy. The present sequence
XX CC is modified clostridial heavy chain-superoxide dismutase conjugate.
XX CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
XX CC translocation domain from diphtheria neurotoxin and a neuronal cell-
XX CC specific binding domain from botulinum neurotoxin type F (BoNT/F).
XX SQ Sequence 862 AA:
XX
XX Query Match 100.0%; Score 144; DB 22; Length 862;
XX Best Local Similarity 100.0%; Pred. No. 7.5e-147;
XX Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 VENVYOMISIDSYINKKAFVITNNRNLGNSRIYINGNLIDKSSISNLDIHSNDILFKI 60
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 576 vfnvgmstisdyinkwlfvltltnrlngrslyingnlidkssisnldhvsdnllfki 635
XX
XX QY 61 VGCNDTRYVGIRFKFVDELTGKTEIETLYSDPPDSILKDFMGNYLLYKRRYLLNLRL 120
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 636 vgcndtryvgirfkfvtelgkteleclysdepdpsllkdtwgnyllynkrryllnllr 695
XX
XX QY 121 TDKSTQNSNPLINQGRGYOKP 144
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 696 tdkstqnsnplinqgrgyvqkp 719
XX
XX RESULT 9
XX AAEO7892 standard; Protein: 887 AA.
XX ID AAEO7892;
XX AC AAEO7892;
XX DT 01-NOV-2001 (first entry)
XX DE Modified clostridial heavy chain-superoxide dismutase conjugate #4.

```

XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
KW superoxide dismutase; SOD; diphtheria neurotoxin; human;  
KW botulinum neurotoxin type F; BONT/F.  
XX  
OS Chimeric - Homo sapiens.  
OS Chimeric - Bacillus stearothermophilus.  
OS Chimeric - Corynebacterium diphtheriae.  
OS Chimeric - Clostridium botulinum.  
OS Chimeric - Synthetic.  
XX  
PN MO200158936-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 04-DEC-2000; 2000WO-GB04644.  
XX  
PR 02-DEC-1999; 98GB-0028530.  
PR 07-APR-2000; 2000GB-0008658.  
XX  
XX  
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
PI Shone CC, Sutton JM, Silman N;  
XX  
DR WPI: 2001-514643/56.  
XX  
PT New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -  
XX  
XX  
XX Example 9; Page 42; 50pp; English.  
XX  
XX The invention relates to a non toxic polypeptide, for delivery of a  
XX therapeutic agent to a neuronal cell, which comprises a binding domain  
XX (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
XX as HC) that binds to the neuronal cell and a translocation domain (amino  
XX terminal half of HC, designated as HN), that translocates the therapeutic  
XX agent into the neuronal cell, where the translocation domain is not a HN  
XX domain of a clostridial neurotoxin and is not a fragment or derivative of  
XX a HN domain of a clostridial toxin. Polypeptides of the invention are  
XX useful for the treatment of a disease state associated with neuronal  
XX cells. The polypeptide constructs are useful for delivering therapeutic  
XX substances to neuronal cells. They are useful to treat disorders of the  
XX CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours,  
XX and infection. They are also useful in gene therapy. The present sequence  
XX is modified clostridial heavy chain-superoxide dismutase conjugate.  
XX This conjugate comprises a mitochondrial leader sequence from human  
XX Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,  
XX liker that can be cleaved by thrombin, translocation domain from  
XX diphtheria neurotoxin and a neuronal cell-specific binding domain from  
XX botulinum neurotoxin type F (BONT/F).  
XX  
XX Sequence 887 AA;  
XX

Query Match	100.0%;	Score 144;	DB 22;	Length 887;
Best Local Similarity	100.0%;	Pred. No. 7.7e-147;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	VENYOMISISYIKKMFVTTNNRLGSRYYINGCNLDEKSIISLGGIHVSDNLFETI	60
Db	601	vfhngmsisdsylnkwlftvltlmmrlngsrlylmgnlidexislnlgdihvsdnllfeti	660
Qy	61	VGCNTRVYGVIRFVYFDELCTELETYSDEPDSILKDWGNYLYNKRYYLNLNR	120
Db	661	vgcnltrvygvlrfvfydelctetletysdepdsilkdwgnlylnkryyllnlr	720
Qy	121	TDKSTIONSNFLINQOQGVYOKP	144
Db	721	tdksitqnsfnlinqgrvgykpr	744

RESULT 10  
 AAE07901  
 ID AAE07901 standard; Protein; 1032 AA.  
 AC AAE07901;  
 XX  
 DT 01-NOV-2001 (first entry)  
 XX  
 DE C. botulinum C2 translocation domain with BoNT/F-binding domain #2.  
 XX  
 KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
 KW tumour; infection; neurodegenerative disease; gene therapy;  
 KW botulinum neurotoxin type F; BoNT/F.  
 XX  
 OS Clostridium botulinum.  
 OS  
 PN WO200158936-A2.  
 PN  
 PD 16-AUG-2001.  
 PD  
 PF 04-DEC-2000; 2000WO-GB04644.  
 PF  
 XX  
 PR 02-DEC-1999; 99GB-0028530.  
 PR 07-APR-2000; 2000GB-0008658.  
 PA  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 XX  
 PI Shone CC, Sutton JM, Silman N;  
 XX  
 DR WPI; 2001-514643/56.  
 XX  
 PT New non toxic polypeptide for delivery of a therapeutic agent for the  
 PT treatment of a CNS disorder comprising a binding domain that  
 PT translocates the therapeutic agent into the neuronal cells -  
 XX  
 PS Example 2; Page 48; 50pp; English.  
 XX  
 CC The invention relates to a non toxic polypeptide, for delivery of a  
 CC therapeutic agent to a neuronal cell, which comprises a binding domain  
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
 CC as HC) that binds to the neuronal cell and a translocation domain (amino  
 CC terminal half of HC, designated as HN), that translocates the therapeutic  
 CC agent into the neuronal cell, where the translocation domain is not a HN  
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
 CC useful for the treatment of a disease state associated with neuronal  
 CC cells. The polypeptide constructs are useful for delivering therapeutic  
 CC substances to neuronal cells. They are useful to treat disorders of the  
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
 CC and infection. They are also useful in gene therapy. The present sequence  
 CC is C. botulinum C2 enterotoxin translocation domain with botulinum  
 CC neurotoxin type F (BoNT/F) binding domain used in the exemplification of  
 CC the invention.  
 XX  
 SQ Sequence 1032 AA;

Query Match	100.0%;	Score 144;	DB 22;	Length 1032;
Best Local Similarity	100.0%;	Pred. No. 8.9e-147;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy	1	VETRYOMISIDYINKKATFEVITNNRNGNSITYINGNIDKSTSNJGDHVSQNIJEKI	60
Oy	746	vfnymqmslsdylnkwlfvctlmnrlgnrsllyngnlldeksismjgdhvsnsnllfkl	805
Oy	61	VGQNDTRVYGRYRFRVPTTELKGTETETELYSDEEDPSILKDDPMWNYLLNKKRYLLMLR	120
Oy	806	vqcnclrvlgvrfvrlvcltelgketelelysdoposllkafwngyllnkrxylmlrl	865
Oy	121	TDKSITQNSNPLNMQRGVYQKP	144
Oy	866	tdksitqnsnplnlnqgrgvqkp	889

```
RESULT 11
AA93309 ID AA93309 standard; protein: 1059 AA.
XX
AC AA93309;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KM Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KM neuronal cell targeting component; NCTC; neuronal disease;
KM oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KM Huntington's disease; motor neurone disease;
KM botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Bacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN MO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB03699.
XX
PR 05-NOV-1998; 98GB-0024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI: 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a
PT cleavable linker to a neuronal cell targeting component useful for
PT delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
XX Disclosure: Page 48-51; 65pp; English.
XX
XX The present sequence represents a construct of the invention, comprising
XX a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
XX can be cleaved by thrombin, and a heavy chain derived from botulinum
XX neurotoxin serotype F. The specification describes a composition for
XX delivery of SOD to neuronal cells. The composition comprises SOD linked,
XX by a cleavable linker, to a neuronal cell targeting component (NCTC).
XX This component has a domain that binds to a neuronal cell and a
XX domain that translocates the SOD of the composition into the neuronal
XX cell. After translocation, the linker is cleaved to release the SOD.
XX The composition is useful for treating neuronal diseases caused or
XX augmented by oxidative stress, such as ischemic stroke, trauma,
XX Parkinson's disease, Huntington's disease and motor neurone diseases.
XX
SQ Sequence 1059 AA:
Query Match 100.0%; Score 144; DB 21; Length 1059;
Best Local Similarity 100.0%; Pred. No. 9.1e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VFNVTQMISIDYINKMIFVTITNNRLGNSRIYINGNLIDKRSISNLGDIHVSNDILFKI 60
Db 773 vfnvtqmisisdyinkmifvtitnnrlgnsriyngnlidkrsisnlgdihvsndilfk1 832
QY 61 VGCNDTRVYGIIRFKVFDTELKTEIETLYSDPPSILKDFMGNYLLVNRKRYLLNLRL 120
Db 833 vgcndtrvygirfkvfdtelkteietlysdppsilkdffwgnyllnkrkryllnlllr 892
QY 121 TDKSTQNSNLFNLINQRCGVQKP 144
Db 893 tdkstqnsnflnlnqrcgvqkp 916
```

```
RESULT 12
AA93312 ID AA93312 standard; protein: 1084 AA.
XX
AC AA93312;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX
KM Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KM neuronal cell targeting component; NCTC; neuronal disease;
KM oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KM Huntington's disease; motor neurone disease;
KM botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Homo sapiens.
OS Bacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN MO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB03699.
XX
PR 05-NOV-1998; 98GB-0024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI: 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a
PT cleavable linker to a neuronal cell targeting component useful for
PT delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
XX Disclosure: Page 57-60; 65pp; English.
XX
XX The present sequence represents a construct of the invention, comprising
XX a mitochondrial leader sequence from human manganese superoxide
XX dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker
XX that can be cleaved by thrombin, and a heavy chain derived from
XX botulinum neurotoxin serotype F. The specification describes a
XX composition for delivery of SOD to neuronal cells. The composition
XX comprises SOD linked, by a cleavable linker, to a neuronal cell
XX targeting component (NCTC). This component has a domain that binds
XX to a neuronal cell and a domain that translocates the SOD of the
XX composition into the neuronal cell. After translocation, the linker
XX is cleaved to release the SOD. The composition is useful for treating
XX neuronal diseases caused or augmented by oxidative stress, such as
XX ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
XX motor neurone diseases.
XX
SQ Sequence 1084 AA:
Query Match 100.0%; Score 144; DB 21; Length 1084;
Best Local Similarity 100.0%; Pred. No. 9.3e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VFNVTQMISIDYINKMIFVTITNNRLGNSRIYINGNLIDKRSISNLGDIHVSNDILFKI 60
Db 798 vfnvtqmisisdyinkmifvtitnnrlgnsriyngnlidkrsisnlgdihvsndilfk1 857
QY 61 VGCNDTRVYGIIRFKVFDTELKTEIETLYSDPPSILKDFMGNYLLVNRKRYLLNLRL 120
Db 858 vgcndtrvygirfkvfdtelkteietlysdppsilkdffwgnyllnkrkryllnlllr 917
QY 121 TDKSTQNSNLFNLINQRCGVQKP 144
```

Db 918 tdkstqnsfllnngqrgvygkp 941

## RESULT 13

AAE07900 AAE07900 standard; Protein; 1092 AA.

AC AAE07900;

DT 01-NOV-2001 (first entry)

DE C. botulinum C2 translocation domain with BONT/F-binding domain #1.

KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;

KW tumour; infection; neurodegenerative disease; gene therapy;

KM botulinum neurotoxin type F; BONT/F.

OS Clostridium botulinum.

PN W0200158936-A2.

PD 16-AUG-2001.

PF 04-DEC-2000; 2000WO-GB04644.

PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Shone CC, Sutton JM, Silman N;

DR WPI; 2001-514643/56.

XX New non toxic polypeptide for delivery of a therapeutic agent for the

PT treatment of a CNS disorder comprising a binding domain that

PT translocates the therapeutic agent into the neuronal cells -

XX Example 2; Page 47; 50pp; English.

CC The invention relates to a non toxic polypeptide, for delivery of a

CC therapeutic agent to a neuronal cell, which comprises a binding domain

CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated

CC as Hc) that binds to the neuronal cell and a translocation domain (amino

CC terminal half of HC, designated as HN), that translocates the therapeutic

CC agent into the neuronal cell, where the translocation domain is not a HN

CC domain of a clostridial neurotoxin and is not a fragment or derivative of

CC a HN domain of a clostridial toxin. Polypeptides of the invention are

CC useful for the treatment of a disease state associated with neuronal

CC cells. The polypeptide constructs are useful for delivering therapeutic

CC substances to neuronal cells. They are useful to treat disorders of the

CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours

CC and infection. They are also useful in gene therapy. The present sequence

CC is C. botulinum C2 enterotoxin translocation domain with botulinum

CC neurotoxin type F (BONT/F) binding domain used in the exemplification of

CC the invention.

CC Sequence 1092 AA;

SQ Query Match 100.0%; Score 144; DB 22; Length 1092;

Best Local Similarity 100.0%; Pred. No. 9.4e-147;

Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VENTOMISISDYINKKIFVTITNNRLGNSRIYINGNIDKSTSNCGDIHVSNDILFKI 60

Db 806 vfnlytqmstsdylnkwlftltnrignsriyngnldkstsngdihvsdnllfki 865

QY 61 VGCNDTRYVGIIRYFKVDTDELGKTEITLXSDPEPDSILKDFMGNLYLLNKRYYLNLRL 120

Db 866 vgcndtryvgiirfkdvtelgkietlxsdepdpssllkdfwgnlyllnkrYYLnlrlr 925

QY 121 TDKSTQNSFLLNNGQRGVYQKP 144

Db 926 tdkstqnsfllnngqrgvygkp 949

## RESULT 14

AAE07898 AAE07898 standard; Protein; 660 AA.

AC AAE07898;

DT 01-NOV-2001 (first entry)

DE Modified clostridial heavy chain fragment #5.

KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;

KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;

KM diphtheria neurotoxin; tetanus neurotoxin; TeNT;

KW botulinum neurotoxin type F; BONT/F.

OS Chimeric - Corynebacterium diphtheriae.

OS Chimeric - Clostridium tetani.

PN W0200158936-A2.

PD 16-AUG-2001.

PF 04-DEC-2000; 2000WO-GB04644.

PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Shone CC, Sutton JM, Silman N;

DR WPI; 2001-514643/56.

XX New non toxic polypeptide for delivery of a therapeutic agent for the

PT treatment of a CNS disorder comprising a binding domain that

PT translocates the therapeutic agent into the neuronal cells -

XX Example 2; Page 46; 50pp; English.

CC The invention relates to a non toxic polypeptide, for delivery of a

CC therapeutic agent to a neuronal cell, which comprises a binding domain

CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated

CC as Hc) that binds to the neuronal cell and a translocation domain (amino

CC terminal half of HC, designated as HN), that translocates the therapeutic

CC agent into the neuronal cell, where the translocation domain is not a HN

CC domain of a clostridial neurotoxin and is not a fragment or derivative of

CC a HN domain of a clostridial toxin. Polypeptides of the invention are

CC useful for the treatment of a disease state associated with neuronal

CC cells. The polypeptide constructs are useful for delivering therapeutic

CC substances to neuronal cells. They are useful to treat disorders of the

CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours

CC and infection. They are also useful in gene therapy. The present sequence

CC is modified clostridial heavy chain fragment. This sequence is

CC constructed by fusing the binding domain which is a hybrid of botulinum

CC neurotoxin type F (BONT/F) and tetanus neurotoxin (TeNT) domain II with

CC translocation domain of diphtheria neurotoxin.

SQ Sequence 660 AA;

Query Match 47.9%; Score 69; DB 22; Length 660;

Best Local Similarity 100.0%; Pred. No. 7.1e-66;

Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VENTOMISISDYINKKIFVTITNNRLGNSRIYINGNIDKSTSNCGDIHVSNDILFKI 60

Db 359 vfnlytqmstsdylnkwlftltnrignsriyngnldkstsngdihvsdnllfki 418



OY 61 VGCNDRYV 69  
 Db 419 vgcndryv 427

Search completed: August 15, 2002, 11:12:27  
 JDB Time: 319 sec

RESULT 15

AA068399

ID AA068399 standard; Protein: 448 AA.

AC AA068399;

DF 07-DEC-1998 (first entry)

DE Clostridium botulinum type F toxin C fragment.

KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;  
 botulism; BoLF.

OS Clostridium botulinum serotype F strain 202F (ATCC 23387).  
 OS Synthetic.

Key Location/Qualifiers

FT Peptide 1..21 /note= "N-terminal His tag"

PN WO9808540-A1.

PD 05-MAR-1998.

PE 28-AUG-1997; 97WO-US15394.

PR 28-AUG-1996; 96US-0704159.

PA (OPHI-) OPHIDIAN PHARM INC.

PI Thalley BS, Williams JA;

DR WPI: 1998-230234/20.

DR N-PSDB: AAV30593.

XX Host cell containing recombinant expression vector encoding

PT Clostridium botulinum type B or E toxin - useful to treat humans

PS Example 48: Page 364-365; 428bp; English.

CC This is the amino acid sequence of the histidine-tagged C fragment  
 of Clostridium botulinum (202F strain) type F neurotoxin, encoded  
 by a DNA sequence (see AAV30593) in plasmid pETHisD. This vector  
 can be used to express BotC soluble C fragment in Escherichia  
 coli host cells, with the recombinant C fragment being purified on  
 an affinity column. The invention relates to recombinant proteins  
 derived from C. botulinum toxins, especially type B and type F  
 toxins. Methods are provided which allow for the isolation of  
 CC soluble recombinant proteins free of significant endotoxin  
 contamination. Preferred hosts for production of recombinant  
 CC proteins are E. coli, insect cells and yeast cells. The  
 CC recombinant toxins are used as immunogens for the production of  
 CC vaccines and antitoxins that are useful in the treatment of humans  
 CC and animals at risk of intoxication with clostridial toxin.

XX Sequence 448 AA;

Query Match 18.8%; Score 27; DB 19; Length 448;

Best Local Similarity 100.0%; Pred. No. 1.3e-20;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 YINKWIEVTITNRLGNSRIYINGNLI 39  
 Db 180 yinkwiefvtitnrlgnsriyingnli 206

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:14:06 ; Search time 47.36 Seconds

(without alignments)  
292.164 Million cell updates/sec

Title: US-08-981-087a-3

Sequence: 1 VENTYOMISIDYINKMFV.....ITONSFNLINQGVYQKP 144

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 28338 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 28338

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR-71:\*\*

1: PIR:\*\*  
2: PIR:\*\*  
3: PIR:\*\*  
4: PIR:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	18.8	1274	2 I40813	neurotoxin type F
2	22	15.3	1268	2 S33411	botulinum neurotoxin
3	15	10.4	1251	2 JH0256	botulinum neurotoxin
4	15	10.4	1252	2 S21178	botulinum neurotoxin
5	11	7.6	1286	1 BPC1AB	botulinum neurotoxin (EC 3)
6	11	7.6	1296	2 I40645	botulinum neurotoxin
7	9	6.2	366	2 S48110	neurotoxin type F
8	8	5.0	540	2 B97350	hypothetical prote
9	8	5.0	1297	2 S37791	neurotoxin - Clostr
10	8	4.9	136	2 S20950	myosin regulatory
11	7	4.9	200	2 A81285	probable membrane
12	7	4.9	202	2 S20952	myosin regulatory
13	7	4.9	204	2 S22715	myosin regulatory
14	7	4.9	226	2 A12668	hypothetical prote
15	7	4.9	231	2 S28494	dx protein - Clo
16	7	4.9	243	2 A91044	hypothetical prote
17	7	4.9	243	2 D85888	hypothetical prote
18	7	4.9	265	2 AG50814	cob(1)alamin adeno
19	7	4.9	267	2 G97450	ABC transporter, A
20	7	4.9	267	2 F81029	type II restrictio
21	7	4.9	307	2 B65021	cob(1)alamin adeno
22	7	4.9	307	2 C83188	probable ATP-bind
23	7	4.9	358	2 S07594	hypothetical prote
24	7	4.9	379	2 B69344	hypothetical prote
25	7	4.9	432	2 D64512	hypothetical prote
26	7	4.9	446	2 T31644	hypothetical prote
27	7	4.9	461	2 B97305	probable cation ef
28	7	4.9	508	2 B6525	hypothetical prote
29	7	4.9	608	2 E71859	phosphogluconate d

#### ALIGNMENTS

30	7	4.9	608	2 D64657	phosphogluconate d
31	7	4.9	665	1 VCYDEM	env polyprotein -
32	7	4.9	814	2 B96630	phosphatidylinosit
33	7	4.9	944	2 D82926	hypothetical prote
34	7	4.9	1023	2 T26261	integrin alpha-9-c
35	7	4.9	1035	2 T58409	hypothetical prote
36	7	4.9	1270	2 T21269	botulinum neurotox
37	7	4.9	1276	2 S11455	botoxilysin (EC 3
38	7	4.9	1291	1 A48940	non-proteolytic bo
39	7	4.9	1291	1 A40631	outer membrane pro
40	7	4.9	1651	2 JC1340	hypothetical prote
41	7	4.9	1655	2 E97835	hypothetical prote
42	6	4.2	30	2 G95031	protein F18014.13
43	6	4.2	35	2 B86327	H+-transporting tw
44	6	4.2	48	1 EWA58	H+-transporting tw
45	6	4.2	48	1 EWA58M	H+-transporting tw

#### RESULT

neurotoxin type F - Clostridium botulinum

C.Species: Clostridium botulinum

C.Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999

C.Accession: I40813; S48108

R.East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, FEMS Microbiol. Lett. 96, 225-230, 1992

A>Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.

A.Reference number: I40644

A.Accession: I40813

A>Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-1274 <RES>

A.Cross-references: GB:M92906; NID:q144866; PIDN:AAA23263.1; PID:q144867

R.Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A>Title: Gene probes for identification of the botulinum neurotoxin gene and specific

A.Reference number: S48103; MUID:94013372

A.Accession: S48108

A>Status: preliminary; translation not shown

A.Molecule type: DNA

A.Residues: 634-1002 <CAN>

A.Cross-references: EMBL:X70816; NID:q407788; PIDN:CAA50147.1; PID:q407789

C.Superfamily: tetanus toxin

C.Keywords: neurotoxin

Query Match 18.8% Score 27; DB 2; Length 1274;

Best Local Similarity 100.0%; Pred No. 4.9e-20;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 YINKMFVITNNRNGNSRIYINGNLI 39

DB 1006 YINKMFVITNNRNGNSRIYINGNLI 1032

#### RESULT

botulinum neurotoxin type F - Clostridium botulinum

C.Species: Clostridium botulinum

C.Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999

C.Accession: S33411; S31860

R.Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P FEMS Microbiol. Lett. 108, 175-182, 1993

A>Title: Nucleotide sequence of the gene coding for Clostridium botulinum type F neuroto

A.Reference number: S33411; MUID:93252228

A.Accession: S33411

A>Status: preliminary

A.Molecule type: DNA

A.Residues: 1-1268 <THO>

A.Cross-references: EMBL:X68262; NID:q49138; PIDN:CAA48329.1; PID:q49139

C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 15.3% Score 22; DB 2; Length 1268;  
Best Local Similarity 100.0%; Pred. No. 9.7e-15;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 DNLEKIVGNDTRYVYIRYK 75  
|||||  
Db 1036 DNLEKIVGNDTRYVYIRYK 1057

## RESULT 3

JH0256  
botulinum neurotoxin type E precursor - Clostridium butyricum  
C:Species: Clostridium butyricum  
C>Date: 30-Jun-1992 #sequence\_revision 15-May-1998 #text\_change 16-Jul-1999

C:Accession: JH0256; S16145  
R:Poulet, S.; Hauser, D.; Quanz, M.; Nlemann, H.; Popoff, M.R.  
Biochem. Biophys. Res. Commun. 183, 107-113, 1992  
A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type  
A:Reference number: JH0256; MUID:92181428  
A:Accession: JH0256

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA  
A:Residues: 1-27, 'E', 29-1251 <POU>

A:Cross-references: EMBL:X62088; NID:940379

A:Experimental source: strains ATCC 43181 and ATCC 43755

R:Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.  
J. Gen. Microbiol. 137, 519-525, 1991

A:Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxin  
A:Reference number: S16145; MUID:91237316  
A:Accession: S16145

A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-229, 'M', 231-252 <FUJ>

A:Cross-references: EMBL:X51180; NID:940407; PIDN:CAA37321.1; PID:940408

A:Experimental source: strain BL6340

C:Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release

C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

F:2-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>  
F:423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>  
F:412-426/Disulfide bonds: #status predicted

Query Match 10.4% Score 15; DB 2; Length 1251;  
Best Local Similarity 100.0%; Pred. No. 2.5e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 ISDYINKWIEVTIN 24  
|||||  
Db 983 ISDYINKWIEVTIN 997

## RESULT 4

S21178  
botulinum neurotoxin type E precursor - Clostridium botulinum

C:Species: Clostridium botulinum  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Oct-1999

C:Accession: S21178; S48107; JH0257; B35294; A60027; S18111

R:Wielan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.  
Eur. J. Biochem. 204, 657-667, 1992

A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin

A:Reference number: S21178; MUID:92174922

A:Accession: S21178

A:Molecule type: DNA  
A:Residues: 1-1252 <WHE>

A:Cross-references: EMBL:X62683; NID:940397; PIDN:CAA44558.1; PID:940398

R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
A:Reference number: S48103; MUID:94013372

A:Accession: S48107

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 616-982 <CAN>

A:Cross-references: EMBL:X70815; NID:9407786; PIDN:CAA50146.1; PID:9407787

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

R:Poulet, S.; Hauser, D.; Quanz, M.; Nlemann, H.; Popoff, M.R.  
Biochem. Biophys. Res. Commun. 183, 107-113, 1992

A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum t

A:Reference number: JH0256; MUID:92181428

A:Accession: JH0257

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA  
A:Residues: 1-176, 'R', 178-197, 'C', 199-339, 'R', 341-772, 'I', 774-962, 'FE', 965-966, 'R', 96

A:Cross-references: EMBL:X62089; NID:940393; PIDN:CAA3999.1; PID:940394

A:Experimental source: strain Beluga

R:Bliz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Nlemann, H.  
J. Biol. Chem. 265, 9153-9158, 1990

A:Title: The complete sequence of botulinum neurotoxin type A and comparison with oth

A:Reference number: A35294; MUID:90264400

A:Accession: B35294

A:Status: not compared with conceptual translation

A:Molecule type: DNA  
A:Residues: 1-176, 'R', 178-252 <BIN>

A:Experimental source: strain Beluga

R:Gimenez, J.A.; Dasgupta, B.R.  
Biochimie 72, 213-217, 1990

A:Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the

A:Reference number: A60027; MUID:90344918

A:Accession: A60027

A:Molecule type: protein

A:Residues: 420-427 <GIN>

A:Experimental source: strain Beluga

A:Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin

C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit

C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

F:1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>  
F:423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>  
F:412-426/Disulfide bonds: #status predicted

Query Match 10.4% Score 15; DB 2; Length 1252;  
Best Local Similarity 100.0%; Pred. No. 2.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 ISDYINKWIEVTIN 24  
|||||  
Db 983 ISDYINKWIEVTIN 997

RESULT 5  
B7CLAB  
bontolixysin (EC 3.4.24.69) A precursor - Clostridium botulinum

N:Alternate names: botulinum neurotoxin type A  
C:Species: Clostridium botulinum

C>Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 18-Jun-1999

C:Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000

R:Bliz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Nlemann, H.  
J. Biol. Chem. 265, 9153-9158, 1990

A:Title: The complete sequence of botulinum neurotoxin type A and comparison with oth

A:Reference number: A35294; MUID:90264400

A:Accession: A35294

A:Molecule type: DNA  
A:Residues: 1-1296 <BIN>

A:Cross-references: GB:M30196; NID:9144864; PIDN:AAA23262.1; PID:9144865

A:Experimental source: strain 62A, subtype A

R:Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson,  
Eur. J. Biochem. 189, 73-81, 1990

A:Title: The complete amino acid sequence of the Clostridium botulinum type A neuroto

A:Reference number: S09492; MUID:90235864  
A:Accession: S09492  
A:Molecule type: DNA  
A:Residues: 1'G',3'-26,'V',28-1296 <THO>  
A:Cross-references: EMBL:X52066; NID:940381; PIDN:CAA36289.1; PID:940382  
A:Experimental source: NCIC 2916  
R:Rajita, K.; Fujinaga, T.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.  
FBS Lett. 376, 41-44, 1995  
A:Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components  
A:Reference number: S67988; MUID:96096783  
A:Accession: S68220  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-12 <FU>  
A:Cross-references: EMBL:D67030; DBJ:D50421; NID:92160224  
R:Beckley, M.J.; Somers, E.; Desgupta, B.R.  
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989  
A:Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term  
A:Reference number: A33401; MUID:89350959  
A:Accession: A33401  
A:Molecule type: DNA  
A:Residues: 1-35 <BE>  
A:Cross-references: GB:M27892; NID:9144880; PIDN:AAA23269.1; PID:9551776  
R:Gimenez, J.A.; Dasgupta, B.R.  
J. Protein Chem. 12, 351-363, 1993  
A:Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, at  
A:Reference number: A53884; MUID:94000342  
A:Accession: A53884  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 867-880;1148-1217,'Y',1219 <GIM>  
A:Experimental source: strain Hall  
R:Dasgupta, B.R.; Dekleva, M.L.  
Biochimie 72, 661-664, 1990  
A:Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and aro  
A:Reference number: A60025; MUID:91120847  
A:Accession: A60025  
A:Molecule type: protein  
A:Residues: 2-6;445-453,'X',455-457 <DAS1>  
R:Dasgupta, B.R.; Foley, J.; Niece, R.  
Biochemistry 26, 4162, 1987  
A:Title: Partial sequence of the light chain of botulinum neurotoxin type A.  
A:Reference number: A27000  
A:Accession: A27000  
A:Molecule type: protein  
A:Residues: 2-47 <DAS2>  
R:Binz, T.; Blaszi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;  
J. Biol. Chem. 269, 1617-1620, 1994  
A:Title: Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.  
A:Reference number: A49708; MUID:94124495  
A:Contents: annotation  
C:Comment: botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap  
C:Genetics:  
A:Gene: atx; botA  
C:Function:  
A:Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal associate  
C:Superfamily: tetanus toxin  
C:Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot  
F:244/Produce: botulinum A light chain #status experimental <IGH>  
F:445/Produce: botulinum A heavy chain #status experimental <HVT>  
F:223/Binding site: zinc (His) #status predicted  
F:224/Active site: Glu #status predicted

Query Match 7.6%; Score 11; DB 1; Length 1296;  
Best Local Similarity 100.0%; Pred. No. 0.0044;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 WIFVITNNRL 27  
|||||  
Db 1014 WIFVITNNRL 1024

RESULT 6  
140645  
botulinum neurotoxin type A - Clostridium botulinum  
C:Species: Clostridium botulinum  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 16-Jul-1999  
C:Accession: 140645  
R:Williams, A.; East, A.K.; Lawson, P.A.; Collins, M.D.  
Res. Microbiol. 144, 547-556, 1993  
A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type  
A:Reference number: 140645; MUID:94143603  
A:Accession: 140645  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1296 <RES>  
A:Cross-references: EMBL:X73423; NID:9507070; PIDN:CAA51824.1; PID:9507071  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 7.6%; Score 11; DB 2; Length 1296;  
Best Local Similarity 100.0%; Pred. No. 0.0044;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 WIFVITNNRL 27  
|||||  
Db 1014 WIFVITNNRL 1024

RESULT 7  
548110  
neurotoxin type F - Clostridium botulinum (fragment)  
C:Species: Clostridium botulinum  
C:Date: 14-Jul-1995 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: 548110  
R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
A:Reference number: 548103; MUID:94013372  
A:Accession: 548110  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-366 <CAM>  
A:Cross-references: EMBL:X70821; NID:9407792; PIDN:CAA50152.1; PID:9407793  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 6.2%; Score 9; DB 2; Length 366;  
Best Local Similarity 100.0%; Pred. No. 0.19;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VENTYOMIS 9  
|||||  
Db 358 VENTYOMIS 366

RESULT 8  
B87350  
hypothetical protein CC0813 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: B87350  
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: B87350  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-540 <STO>

A;Cross-references: GB:AE005673; NID:913422062; PIDN:AAK22798.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC0813

Query Match 5.6%; Score 8; DB 2; Length 540;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 PDPSTLKD 101  
|||||  
Db 71 PDPSTLKD 78

RESULT 9  
S39791  
neurotoxin - Clostridium botulinum  
C;Species: Clostridium botulinum  
C;Date: 07-Oct-1994 #sequence\_revision 01-Dec-1995 #text\_change 16-Jul-1999  
C;Accession: S39791  
R;Campbell, K.; Collins, M.D.; East, A.K.;  
Biochim. Biophys. Acta 1216, 487-491, 1993  
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a  
A;Reference number: S39791; MUID:94092745  
A;Accession: S39791  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1297 <CAM>  
A;Cross-references: EMBL:X74162; NID:9441275; PIDN:CAA52275.1; PID:9441276  
C;Superfamily: tetanus toxin  
C;Keywords: neurotoxin

Query Match 5.6%; Score 8; DB 2; Length 1297;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ISDYINKW 17  
|||||  
Db 1002 ISDYINKW 1009

RESULT 10  
S20990  
myosin regulatory light chain 2 - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 16-Jul-1999  
C;Accession: S20990  
R;Alt, F.W.  
submitted to the EMBL Data Library, May 1992  
A;Reference number: S20990  
A;Accession: S20990  
A;Molecule type: mRNA  
A;Residues: 1-156 <ALT>  
A;Cross-references: EMBL:X65979  
C;Superfamily: calmodulin; calmodulin repeat homology  
C;Keywords: calcium binding; EF hand  
F;14-46/Domain: calmodulin repeat homology <EF1>  
F;84-116/Domain: calmodulin repeat homology <EF2>

Query Match 4.9%; Score 7; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 FKVPDTE 80  
|||||  
Db 93 FKVPDTE 99

RESULT 11  
A81295  
probable membrane protein Cj1484c [Imported] - Campylobacter jejuni (strain NCTC 11168)

C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 28-Jul-2000  
C;Accession: A81295  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chill  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
A;Reference number: A81295; MUID:20150912  
A;Accession: A81295  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-200 <PAR>  
A;Cross-references: GB:AL139078; GB:AL111168; NID:96968723; PIDN:CAB73906.1; PID:9696  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj1484c  
C;Superfamily: Campylobacter jejuni probable membrane protein Cj1484c

Query Match 4.9%; Score 7; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 KTEJETL 89  
|||||  
Db 66 KTEJETL 72

RESULT 12  
S20992  
myosin regulatory light chain 2 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 13-Aug-1999  
C;Accession: S20992  
R;Alt, F.W.  
submitted to the EMBL Data Library, May 1992  
A;Reference number: S20990  
A;Accession: S20992  
A;Molecule type: DNA  
A;Residues: 1-202 <ALT>  
A;Cross-references: EMBL:X65981; NID:953747; PIDN:CAA6796.1; PID:953748  
C;Genetics:  
A;Introns: 128/1  
C;Superfamily: calmodulin; calmodulin repeat homology  
C;Keywords: calcium binding; EF hand  
F;60-92/Domain: calmodulin repeat homology <EF1>  
F;130-162/Domain: calmodulin repeat homology <EF2>

Query Match 4.9%; Score 7; DB 2; Length 202;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 FKVPDTE 80  
|||||  
Db 139 FKVPDTE 145

RESULT 13  
S22715  
myosin regulatory light chain 2 - human  
N;Alternate names: pre-lymphocyte-specific regulatory light chain PLR1C  
C;Species: Homo sapiens (man)  
C;Date: 03-May-1994 #sequence\_revision 20-Feb-1995 #text\_change 16-Jul-1999  
C;Accession: S22715  
R;Oltz, E.M.; Yancopoulos, G.D.; Morrow, M.A.; Rolink, A.; Lee, G.; Wong, F.; Kaplan,  
EMBO J. 11, 2759-2767, 1992  
A;Title: A novel regulatory myosin light chain gene distinguishes pre-B cell subsets  
A;Reference number: S22715; MUID:92331628  
A;Accession: S22715  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-204 <OLT>

C:Superfamily: calmodulin; calmodulin repeat homology  
 C:Keywords: calcium binding; EF hand  
 F:62-94/Domain: calmodulin repeat homology <EF1>  
 F:132-164/Domain: calmodulin repeat homology <EF2>

Query Match 4.9%; Score 7; DB 2; Length 204;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 FKVPDTE 80  
 |||||  
 Db 141 FKVPDTE 147

RESULT 14

AI2668  
 Hypothetical protein Atu0751 [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C:Accession: AI2668  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AI2668  
 A>Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-226 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AAI41767.1; PID:g17739119; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu0751  
 A:Map position: circular chromosome

Query Match 4.9%; Score 7; DB 2; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LKTEIE 87  
 |||||  
 Db 105 LKTEIE 111

RESULT 15

S28494  
 dtxA protein - Clostridium difficile  
 N:Alternate names: tcdC  
 C:Species: Clostridium difficile  
 C>Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 15-Oct-1999  
 C:Accession: J05344; S28494  
 R:Brann, V.; Hunsberger, T.; Leukel, P.; Sauerborn, M.; von Eichel-Streiber, C.  
 Gene 181, 29-38, 1996  
 A:Title: Definition of the single integration site of the pathogenicity locus in Clostr  
 A:Reference number: J05340; PMID:97128764  
 A:Accession: J05344  
 A:Molecule type: DNA  
 A:Residues: 1-231 <BR>  
 A:Cross-references: EMBL:X92982; NID:g1770128; PIDN:CMA63565.1; PID:e212238; PID:g177013  
 A:Experimental source: strain VP10463  
 C:Genetics:  
 A:Gene: dtxA; tcdC

Query Match 4.9%; Score 7; DB 2; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KTEIETL 89  
 |||||  
 Db 70 KTEIETL 76

Search completed: August 15, 2002, 11:14:07  
 Job time: 259 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:38 ; Search time 24.69 Seconds

(without alignments)  
225.825 Million cell updates/sec

Title: US-08-981-087a-3

Perfect score: 144  
Sequence: I VFNNTOMISIDYINKWIFV.....ITQNSFNLINQGRGYQKRP 144

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	18.8	1274	1	BXE_CLOBO
2	13	10.4	1230	1	BXE_CLOBO
3	13	10.4	1230	1	BXE_CLOBO
4	11	7.6	1295	1	BXA2_CLOBO
5	11	7.6	1295	1	BXA2_CLOBO
6	8	5.6	1296	1	BXA2_CLOBO
7	7	4.9	199	1	TDX2_THDAC
8	7	4.9	267	1	EUTT_ECOLI
9	7	4.9	267	1	EUTT_ECOLI
10	7	4.9	358	1	VAL1_CLOV
11	7	4.9	358	1	VAL1_CLOV
12	7	4.9	432	1	Y221_MERJA
13	7	4.9	508	1	Y221_MERJA
14	7	4.9	608	1	EDD_HELPJ
15	7	4.9	608	1	EDD_HELPJ
16	7	4.9	665	1	ENV_MLYMO
17	7	4.9	814	1	PI3K_ARATH
18	7	4.9	942	1	AMPN_MANSE
19	7	4.9	944	1	Y166_UREPA
20	7	4.9	1035	1	IT9A_HUMAN
21	7	4.9	1276	1	BXD_CLOBO
22	7	4.9	1290	1	BXD_CLOBO
23	7	4.9	1550	1	OMP8_RICCN
24	7	4.9	1550	1	OMP8_RICCN
25	6	4.2	48	1	ATP8_RICCN
26	6	4.2	48	1	ATP8_RICCN
27	6	4.2	49	1	ATP8_RICCN
28	6	4.2	49	1	ATP8_RICCN
29	6	4.2	63	1	RI29_BUCAR
30	6	4.2	63	1	RI29_BUCAR
31	6	4.2	74	1	YVFE_VACCC
32	6	4.2	75	1	YVFE_VACCC
33	6	4.2	78	1	HSTO_VIBCH
34	6	4.2	78	1	HSTO_VIBCH

34	6	4.2	114	1	Y211_BUCAI	P57307 buchnera ap
35	6	4.2	128	1	MNR8_EVATR	P28093 evasterias
36	6	4.2	137	1	IPPD_PIG	O29277 sus scrofa
37	6	4.2	146	1	AZUR_METPL	O50400 methylolact
38	6	4.2	151	1	YX98_CLOAB	P33651 clostridium
39	6	4.2	163	1	ATPX_OCHNE	O40608 ochrosphaer
40	6	4.2	167	1	YVLA_BAGSU	P40405 bacillus su
41	6	4.2	181	1	GLD2_HORVU	P55235 hordeum vul
42	6	4.2	189	1	YIMC_CABEL	O01901 caenorhabd
43	6	4.2	191	1	YERA_ARATH	O48670 arabidopsis
44	6	4.2	195	1	YERB_ARATH	O48671 arabidopsis
45	6	4.2	200	1	HBP8_RH1AP	O77422 rhinicephal

## ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description
1	BXE_CLOBO	18.8	1274	1	BXE_CLOBO	
2	BXE_CLOBO	10.4	1230	1	BXE_CLOBO	
3	BXE_CLOBO	10.4	1230	1	BXE_CLOBO	
4	BXA2_CLOBO	7.6	1295	1	BXA2_CLOBO	
5	BXA2_CLOBO	7.6	1295	1	BXA2_CLOBO	
6	BXA2_CLOBO	5.6	1296	1	BXA2_CLOBO	
7	TDX2_THDAC	4.9	199	1	TDX2_THDAC	
8	EUTT_ECOLI	4.9	267	1	EUTT_ECOLI	
9	EUTT_ECOLI	4.9	267	1	EUTT_ECOLI	
10	VAL1_CLOV	4.9	358	1	VAL1_CLOV	
11	VAL1_CLOV	4.9	358	1	VAL1_CLOV	
12	Y221_MERJA	4.9	432	1	Y221_MERJA	
13	Y221_MERJA	4.9	508	1	Y221_MERJA	
14	EDD_HELPJ	4.9	608	1	EDD_HELPJ	
15	EDD_HELPJ	4.9	608	1	EDD_HELPJ	
16	ENV_MLYMO	4.9	665	1	ENV_MLYMO	
17	PI3K_ARATH	4.9	814	1	PI3K_ARATH	
18	AMPN_MANSE	4.9	942	1	AMPN_MANSE	
19	Y166_UREPA	4.9	944	1	Y166_UREPA	
20	IT9A_HUMAN	4.9	1035	1	IT9A_HUMAN	
21	BXD_CLOBO	4.9	1276	1	BXD_CLOBO	
22	BXD_CLOBO	4.9	1290	1	BXD_CLOBO	
23	OMP8_RICCN	4.9	1550	1	OMP8_RICCN	
24	OMP8_RICCN	4.9	1550	1	OMP8_RICCN	
25	ATP8_RICCN	4.2	48	1	ATP8_RICCN	
26	ATP8_RICCN	4.2	48	1	ATP8_RICCN	
27	ATP8_RICCN	4.2	49	1	ATP8_RICCN	
28	ATP8_RICCN	4.2	49	1	ATP8_RICCN	
29	RI29_BUCAR	4.2	63	1	RI29_BUCAR	
30	RI29_BUCAR	4.2	63	1	RI29_BUCAR	
31	YVFE_VACCC	4.2	74	1	YVFE_VACCC	
32	YVFE_VACCC	4.2	75	1	YVFE_VACCC	
33	HSTO_VIBCH	4.2	78	1	HSTO_VIBCH	
34	HSTO_VIBCH	4.2	78	1	HSTO_VIBCH	

```

CC detected action on small molecule substrates.
CC -1 SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M92906; AAA23263.1; -
DR EMBL: S73676; AAC60475.1; -
DR EMBL: X70820; CAA50151.1; -
DR EMBL: X70816; CAA50147.1; -
DR HSSP: P10845; 3BTA.
DR MEROPS: M27.002; -
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_MTPeptidse.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT METAL 437 1274 BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT METAL 227 227 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 228 228 BY SIMILARITY.
FT METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 429 445 INTERCHAIN (PROBABLE).
SQ SEQUENCE 1274 AA; 146709 MW; 5B99756A/438B921 CRC64;

Query Match 18.8%; Score 27; DB 1; Length 1274;
Best Local Similarity 100.0%; Pred. No. 4,3e-20;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 YINKWIFVTITNRLGNSRIYINGNLI 39
DB 1006 YINKWIFVTITNRLGNSRIYINGNLI 1032

RESULT 2
BXE_CLOBO
ID BXE_CLOBO STANDARD; PRT; 1250 AA.
AC 000496;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)
DE (Bontoxilysin E).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacilllus/Clostridium group; Clostridiaceae;
OC Clostridium
OX NCBI_TaxId=1491;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BEUGA;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RT botulinum type E (strain Beuga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RN SEQUENCE FROM N.A.

```

```

RX MEDLINE=92174922; PubMed=1541280;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 204:657-667(1992).
RN [3]
RN SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE=90264400; PubMed=2160960;
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).
RN [4]
RN SEQUENCE OF 1-13.
RX MEDLINE=85197963; PubMed=3888113;
RA Schmidt J.J., Sathiyamoorthy V., Dasgupta B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
RN [5]
RN SEQUENCE OF 419-426.
RX MEDLINE=90344918; PubMed=2116911;
RA Gimenez J.A., Dasgupta B.R.;
RT "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT reveals the site trypsin nicks and homology with tetanus
RT neurotoxin.";
RL Biochimie 72:213-217(1990).
RN [6]
RN IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA Benfenati F., Wilson M.C., Montecucco C.;
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS Lett. 335:99-103(1993).
RN [7]
RN IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94124495; PubMed=8294407;
RA Binz T., Blas J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA Jahn R., Niemann H.;
RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
RL J. Biol. Chem. 269:1617-1620(1994).
CC -1 FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-1-ILE-
CC 181 BOND IN SNAP-25.
CC -1 CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO
CC detected action on small molecule substrates.
CC -1 SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62089; CAA43999.1; -
DR EMBL: X62683; CAA44558.1; -
DR PIR: A60027; A60027.

```

```

DR PIR: B35294; B35294.
DR PIR: JH0257; JH0257.
DR PIR: S08575; S08575.
DR PIR: S18111; S18111.
DR PIR: S21178; S21178.
DR HSSP: P10845; 3BTA.
DR MEROPS: M27_002; -.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOLYLISIN.
DR PRODOM: PD001963; Bontoloxlysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
KW INIT_MET 0
FT CHAIN 1 421
FT METAL 422 1250
FT ACT_SITE 211 211
FT METAL 212 212
FT METAL 215 215
FT DISULFID 411 425
FT CONFLICT 176 176
FT CONFLICT 197 197
FT CONFLICT 339 339
FT CONFLICT 772 772
FT CONFLICT 962 963
FT CONFLICT 966 966
FT CONFLICT 1194 1194
SQ SEQUENCE 1250 AA; 143712 MW; D9PC26DDA041B84 CRC64;

Query Match 10.4%; Score 15; DB 1; Length 1250;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ISDYINKMIFVITN 24
DB 982 ISDYINKMIFVITN 996

RESULT 3
BXE_CLOBO STANDARD; PRT; 1250 AA.
ID BXE_CLOBO
AC P30995;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)
DE (Bontoloxlysin E).
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43181;
RX MEDLINE=92181428; PubMed=1543481;
RA Roulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE OF 1-251 FROM N.A.
RC STRAIN-BL6340;
RX MEDLINE=91237316; PubMed=2033376;
RA Yokosawa N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. Gen. Microbiol. 137:519-525(1991).
RN [3]
RP SEQUENCE OF 1-48.

```

```

RC STRAIN-5262;
RA Gimenex J., Foley J., Dasgupta B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT partial sequence and comparison.";
RL FASEB J 2:A1750-A1750(1988)
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT OF THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPePTIDASE.
CC -1- CATALYTIC ACTIVITY: limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62088; CAA43988.1; -.
DR EMBL: X63180; CAA37321.1; -.
DR PIR: JH0258; JH0258.
DR PIR: S16145; S16145.
DR HSSP: P10845; 3BTA.
DR MEROPS: M27_002; -.
DR InterPro: IPR000395; Bontoloxlysin.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOLYLISIN.
DR PRODOM: PD001963; Bontoloxlysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0
FT CHAIN 1 421
FT CHAIN 422 1250
FT METAL 211 211
FT METAL 212 212
FT ACT_SITE 215 215
FT METAL 411 425
FT DISULFID 176 176
FT CONFLICT 229 229
SQ SEQUENCE 1250 AA; 143265 MW; 8171B5B2C312857 CRC64;

Query Match 10.4%; Score 15; DB 1; Length 1250;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ISDYINKMIFVITN 24
DB 982 ISDYINKMIFVITN 996

RESULT 4
BXAL_CLOBO STANDARD; PRT; 1295 AA.
ID BXAL_CLOBO
AC P10845; P18639; P01561;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
DE (Bontoloxlysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-

```

DE chain; Botulinum neurotoxin A, heavy-chain].  
GN BOTA OR BNA OR ATX.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NCTC 2916;  
RX MEDLINE=90235864; PubMed=2185020;  
RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,  
RT Shone C.C., Atkinson T., Melling J., Minton N.P.;  
RT "The complete amino acid sequence of the Clostridium botulinum type A  
RT neurotoxin, deduced by nucleotide sequence analysis of the encoding  
RT gene.";  
RL Eur. J. Biochem. 189:73-81(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-62A;  
RX MEDLINE=90264400; PubMed=2160960;  
RA Binz B., Kuarazono H., Wille M., Frevent J., Wernars K., Niemann H.;  
RT "The complete sequence of botulinum neurotoxin type A and comparison  
RT with other clostridial neurotoxins.";  
RL J. Biol. Chem. 265:9153-9158(1990).  
RN [3]  
RP SEQUENCE OF 1-65 FROM N.A.  
RC STRAIN-62A;  
RX MEDLINE=97016817; PubMed=8863443;  
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
RT "Organization and phylogenetic interrelationships of genes encoding  
RT components of the botulinum toxin complex in proteolytic Clostridium  
RT botulinum types A, B, and F: evidence of chimeric sequences in the  
RT gene encoding the nontoxic nonhemagglutinin component.";  
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
RN [4]  
RP SEQUENCE OF 1-34 FROM N.A.  
RC STRAIN-HALT;  
RX MEDLINE=89350959; PubMed=2669749;  
RA Beasley M.J., Somers E., Dasgupta B.R.;  
RT "Characterization of botulinum type A neurotoxin gene: delineation of  
RT the N-terminal encoding region.";  
RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).  
RN [5]  
RP SEQUENCE OF 1-18 FROM N.A.  
RC STRAIN-TYPE A NIH;  
RX MEDLINE=96096783; PubMed=8521962;  
RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
RT components of Clostridium botulinum type A progenitor toxins.";  
RL FEBS Lett. 376:41-44(1995).  
RN [6]  
RP SEQUENCE OF 1-16.  
RX MEDLINE=84178501; PubMed=6370252;  
RA Schmidt J.J., Sathymoorthy V., Dasgupta B.R.;  
RT "Partial amino acid sequence of the heavy and light chains of  
RT botulinum neurotoxin type A.";  
RL Biochem. Biophys. Res. Commun. 119:900-904(1984).  
RN [7]  
RP SEQUENCE OF 1-46.  
RA Dasgupta B.R., Foley J., Niece R.;  
RT "Partial sequence of the light chain of botulinum neurotoxin type A.";  
RL Biochemistry 26:4162-4164(1987).  
RN [8]  
RP SEQUENCE OF 1-5 AND 444-456.  
RX MEDLINE=91120847; PubMed=2126206;  
RA Dasgupta B.R., Dekleva M.L.;  
RT "Botulinum neurotoxin type A: sequence of amino acids at the  
RT N-terminus and around the nicking site.";  
RL Biochimie 72:661-664(1990).  
RN [9]  
RP SEQUENCE OF 448-464 AND 872-895.  
RX MEDLINE=89024662; PubMed=3178218;  
RA Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;

RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two  
RT halves and their partial sequences.";  
RL Arch. Biochem. Biophys. 266:142-151(1988).  
RN [10]  
RP SEQUENCE OF 448-482.  
RX MEDLINE=85285016; PubMed=3896784;  
RA Shone C.C., Hambleton P., Melling J.;  
RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin  
RT and purification of two tryptic fragments. Proteolytic action near  
RT the COOH-terminus of the heavy subunit destroys toxin-binding  
RT activity.";  
RL Eur. J. Biochem. 151:75-82(1985).  
RN [11]  
RP IDENTIFICATION OF SUBSTRATE.  
RX MEDLINE=94063091; PubMed=8243676;  
RA Schiavo G., Santucci A., Dasgupta B.R., Menta P.P., Jontes J.,  
RA Benfenati F., Wilson M.C., Montecucco C.;  
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
RT COOH-terminal peptide bonds.";  
RL FEBS Lett. 335:99-103(1993).  
RN [12]  
RP IDENTIFICATION OF SUBSTRATE.  
RX MEDLINE=94124495; PubMed=8294407;  
RA Binz T., Blaszi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
RA Jahn R., Niemann H.;  
RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";  
RL J. Biol. Chem. 269:1617-1620(1994).  
RN [13]  
RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.  
RX PubMed=11700044;  
RA Rigout M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;  
RT "Site-directed mutagenesis identifies active-site residues of the  
RT light chain of botulinum neurotoxin type A.";  
RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).  
RX MEDLINE=98455071; PubMed=9783750;  
RA Lacey D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;  
RT "Crystal structure of botulinum neurotoxin type A and implications  
RT for toxicity.";  
RL Nat. Struct. Biol. 5:898-902(1998).  
CC -1- FUNCTION: Inhibits acetylcholine release. The botulinum toxin  
CC binds with high affinity to peripheral neuronal presynaptic  
CC membrane, is then internalized by receptor-mediated endocytosis.  
CC The C-terminus of the heavy chain (H) is responsible for the  
CC adherence of the toxin to the cell surface while the N-terminus  
CC mediates transport of the light chain from the endocytic vesicle  
CC to the cytosol. After translocation, the light chain (L)  
CC hydrolyzes the 197-Gln-1-Arg-198 bond in SNAP-25, thereby blocking  
CC neurotransmitter release. Inhibition of acetylcholine release  
CC results in flaccid paralysis, with frequent heart or respiratory  
CC failure.  
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO  
CC detected action on small molecule substrates.  
CC -1- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a  
CC heavy chain (H).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- PHARMACEUTICAL: Available under the name BOTOX(R) (Allergan) for  
CC the treatment of strabismus and blepharospasm associated with  
CC dystonia and cervical dystonia. Also used for the treatment of  
CC hemifacial spasm and a number of other neurological disorders  
CC characterized by abnormal muscle contraction.  
CC -1- MISCELLANEOUS: There are seven antigenically distinct forms of  
CC botulinum neurotoxin: Types A, B, C1, D, E, F, and G.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: X52066; CAA36289.1; -
DR EMBL: M30196; AAA23262.1; -
DR EMBL: X92973; CAA63551.1; -
DR EMBL: D67030; BAA11051.1; -
DR EMBL: M27892; AAA23269.1; -
DR PIR: A35284; BRCIAB.
DR PIR: S09492; S09492.
DR PDB: 3BTA; 01-OCT-99.
DR MEROPS: M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_Mpeptidase.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; Bontoxilysin.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW Pharmaceutical; 3D-structure.
FT INIT_MET 0 0
FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT METAL 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT ACT_SITE 222 222 ZINC (CATALYTIC).
FT METAL 223 223 ZINC (CATALYTIC).
FT METAL 226 226 ZINC (CATALYTIC).
FT METAL 261 261 ZINC (CATALYTIC).
FT DISULFID 429 453 INTERCHAIN.
FT DISULFID 1234 1279 INTERCHAIN.
FT TRANSMEM 626 646 POTENTIAL.
FT TRANSMEM 655 675 POTENTIAL.
FT VARIANT 26 26 V->A.
FT MUTAGEN 261 261 E->A: DRASTIC DECREASE IN ENZYMIC ACTIVITY.
FT MUTAGEN 265 265 F->A: DECREASE IN ENZYMIC ACTIVITY.
FT CONFLICT 365 365 Y->A: DECREASE IN ENZYMIC ACTIVITY.
FT CONFLICT 475 476 P->Q (IN REF. 1).
FT CONFLICT 875 875 E->P (IN REF. 9).
FT CONFLICT 891 891 T->L (IN REF. 8).
FT CONFLICT 891 891 S->K (IN REF. 8).
SQ SEQUENCE 1295 AA; 149322 MW; 856342F754862579 CRC64;

```

Query Match 7.6%; Score 11; DB 1; Length 1295;  
Best Local Similarity 100.0%; Pred. No. 0.0021;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 17 WIFVTTNNRL 27  
DB 1013 WIFVTTNNRL 1023

```

RESULT 5
BXA2_CLOBO STANDARD; PRT: 1295 AA.
AC 045894; P77780;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)
DE (Bontoxilysin A) (BOTOX) [contains: Botulinum neurotoxin A, light-
DE chain; Botulinum neurotoxin A, heavy-chain].
GN BOTA OR BMA OR ATX.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1491;
RP 11;
RC STRAIN-Kyoto-F;
RX MEDLINE=94143603; PubMed=8310180;
RA Williams A., East A.K., Lawson P.A., Collins M.D.;
RT Sequence of the gene coding for the neurotoxin of Clostridium
RT botulinum type A associated with infant Botulism: comparison with

```

```

RT other clostridial neurotoxins.";
RN Res. Microbiol. 144:547-556(1993).
RL [2]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN-Kyoto-F;
RX MEDLINE=97016817; PubMed=8863443;
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the non-toxic nonhemagglutinin component.";
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC -I- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC binds with high affinity to peripheral neuronal presynaptic
CC membrane, is then internalized by receptor-mediated endocytosis.
CC The C-terminus of the heavy chain (H) is responsible for the
CC adherence of the toxin to the cell surface while the N-terminus
CC mediates transport of the light chain from the endocytic vesicle
CC to the cytosol. After translocation, the light chain (L)
CC hydrolyzes the 197-Gln-1-Arg-198 bond in SNAP-25, thereby blocking
CC neurotransmitter release. Inhibition of acetylcholine release
CC results in flaccid paralysis, with frequent heart or respiratory
CC failure (by similarity).
CC -I- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO
CC detected action on small molecule substrates.
CC -I- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC heavy chain (H) (By similarity).
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- MISCELLANEOUS: There are seven antigenically distinct forms of
CC botulinum neurotoxin: Types A, B, C, D, E, F, and G.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: X73423; CAA51824.1; -.
DR EMBL: X87974; CAA61234.1; -.
DR HSSP: P10845; 3BTA.
DR MEROPS: M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; Bontoxilysin.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0 0
FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 223 223 BY SIMILARITY.
FT METAL 226 226 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 429 453 INTERCHAIN (BY SIMILARITY).
FT DISULFID 1234 1279 BY SIMILARITY.
FT TRANSMEM 626 646 POTENTIAL.
FT TRANSMEM 655 675 POTENTIAL.
SQ SEQUENCE 1295 AA; 149279 MW; SDA04A13D98D6372 CRC64;

```

Query Match 7.6%; Score 11; DB 1; Length 1295;  
Best Local Similarity 100.0%; Pred. No. 0.0021;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 WIFVTTNNRL 27  
DB 1013 WIFVTTNNRL 1023

```

RESULT 6
BXG_CLOBO ID BXG_CLOBO STANDARD: PRT: 1296 AA.
AC 060393:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BONT/G)
DE (Bontoxilysin G).
GN BONTG.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=113 / 30;
RX MEDLINE=94092745; Pubmed=8268233;
RA Campbell K., Collins M.D., East A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT with other clostridial neurotoxins.";
RL Blochum, Biophys. Acta 1216:487-491(1993).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted (by similarity).
CC -1- BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74162; CA52275.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KM Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT INTL_MET 0 0
FT CHAIN 1 441
FT CHAIN 442 1296
FT METAL 229 229
FT ACT_SITE 230 230
FT METAL 233 233
FT DISULFID 435 449
SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F65C31 CRC64;

```

Query Match 5.6%; Score 8; DB 1; Length 1296;  
 Best Local Similarity 100.0%; Pred. No. 2.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 10 ISDYINKW 17
Db 1001 ISDYINKW 1008
RESULT 7
ID TX2_THEAC STANDARD: PRT: 199 AA.
AC Q9HJL3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable peroxiredoxin 2.
GN TA0954.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; Pubmed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Fritsman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
CC -1- SIMILARITY: BELONGS TO THE AHPc/TSA FAMILY. TDxH SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445066; CAC12083.1; -.
DR InterPro: IPR000866; AHPc-TSA.
DR Pfam; PF00578; AHPc-TSA; 1.
KW Antioxidant; Complete proteome.
FT ACT_SITE 40 40
SQ SEQUENCE 199 AA; 22581 MW; 4E26162F5D58162 CRC64;

```

Query Match 4.9%; Score 7; DB 1; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 5.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 37 NLIIDEKS 43
Db 104 NLIIDEKS 110
RESULT 8
ID EUTL_ECOLI STANDARD: PRT: 267 AA.
AC P76554;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).
GN EUTL OR B2459.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL12 / MG1655;
RX MEDLINE=97426617; Pubmed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1453-1474(1997).  
 CC -1- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY)  
 CC -1- CATALYTIC ACTIVITY: ATP + cob(II)alamin + H(2)O = phosphate +  
 CC diphosphate + adenosylcobalamin.  
 CC -1- PATHWAY: ETHANOLAMINE UTILIZATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000332; AAC75512.1; -.  
 DR Ecogene; EGI4189; eutT.  
 KW Transferase; Complete proteome.  
 SQ SEQUENCE 267 AA; 30171 MW; E51EDAB528B4FA76 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 7.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 115 LN1LRT 121  
 DB 199 LN1LRT 205

RESULT 9  
 EUTL\_SALTY STANDARD: PRT; 267 AA.  
 ID EUTL\_SALTY  
 AC Q92EV4; 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).  
 GN EUTL OR STM2467.  
 OS *Salmonella typhimurium*.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC *Salmonella*.  
 OC NCBI\_TaxID=602;  
 RX NCBI\_TaxID=602;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-LT2;  
 RC MEDLINE=99395039; PubMed=10464203;  
 RA Kofoid E.C., Rapleye C.A., Stojiljkovic I., Roth J.R.;  
 RT "The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium*  
 RT encodes five homologues of carboxysome shell proteins.";  
 RL J Bacteriol. 181:5317-5329(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LT2 / S6SCI412 / ATCC 700720;  
 RA MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 CC -1- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + cob(II)alamin + H(2)O = phosphate +  
 CC diphosphate + adenosylcobalamin.  
 CC -1- PATHWAY: ETHANOLAMINE UTILIZATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF093749; AAC78114.1; -.  
 DR EMBL; AE008811; AAC21361.1; -.  
 DR Ecogene; SG10636; eutT.  
 KW Transferase; Complete proteome.  
 SQ SEQUENCE 267 AA; 30238 MW; 9502A28BFD84DC9E4 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 7.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 115 LN1LRT 121  
 DB 199 LN1LRT 205

RESULT 10  
 VAL1\_CLVK STANDARD: PRT; 358 AA.  
 ID VAL1\_CLVK  
 AC P14982;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE AL1 protein (40.4 kDa protein).  
 GN A11.  
 OS Cassava latent virus (strain West Kenyan 844).  
 CC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.  
 OC NCBI\_TaxID=10818;  
 RX NCBI\_TaxID=10818;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stanley J., Gay M.R.;  
 RT Nucleotide sequence of cassava latent virus DNA.";  
 RL Nature 301:260-262(1983).  
 CC -1- SIMILARITY: BELONGS TO GEMINIVIRUSES A11 PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J02057; -; NOT ANNOTATED.CDS.  
 DR InterPro; IPR001191; Gemin1\_AL1.  
 DR Pfam; PF00799; Gemin1\_AL1; 1.  
 DR PRINTS; PRO0227; GEMCON1AL1.  
 DR ProDom; PD000736; Gemin1\_AL1; 1.  
 KW ATP-binding.  
 FT NE\_BIND 220 227 ATP (POTENTIAL).  
 SQ SEQUENCE 358 AA; 40346 MW; ED173E753EE92D69 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 358;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 TITNRL 27  
 DB 68 TITNRL 74

RESULT 11  
 VAL1\_CLVK STANDARD: PRT; 358 AA.  
 ID VAL1\_CLVK  
 AC P14972;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)

```

DE 01-JUN-1994 (Rel. 29, last annotation update)
DE A11 protein (40.4 kDa protein).
GN A11.
OS Cassava latent virus (strain Nigerian).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
RN NCBI_TaxID=10819;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174930; PubMed=2308831.
RA Morris B., Coates L., Lowe S., Richardson K., Eddy P.;
RT "Nucleotide sequence of the infectious cloned DNA components of
RT African cassava mosaic virus (Nigerian strain).";
RL Nucleic Acids Res. 18:197-198(1990).
CC -1- SIMILARITY: BELONGS TO GEMINIVIRUSES A11 PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17095; CAA34953.1; -.
DR PIR: S07594; S07594.
DR InterPro: IPR001191; Gemini_AL1.
DR Pfam: PF00799; Gemini_AL1.1.
DR PRINTS: PR00227; GEMCOAT1.1.
DR ProDom: PD000736; Gemini_AL1.1.
KM ATP-binding.
FT NP_BIND 220 227 ATP (POTENTIAL).
SQ SEQUENCE 358 AA; 40435 MW; 1DB16B0CB2D5E2C CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 TITNRL 27
DB 68 TITNRL 74
|||||
RESULT 12
Y221_METUA STANDARD; PRT; 432 AA.
ID Y221_METUA
AC 060281;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJECL21.
GN MJECL21.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L77118; AAC37092.1; -.
DR TIGR: MJECL21; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 432 AA; 51081 MW; DBADF2C5C43A4F90 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 YLLNRLR 120
DB 211 YLLNRLR 217
|||||
RESULT 13
YM05_ARCFU STANDARD; PRT; 508 AA.
ID YM05_ARCFU
AC 028078;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2205.
GN AF2205.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Griest N.H., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000953; AAB89056.1; -.
DR TIGR: AF2205; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 29 POTENTIAL.
SQ SEQUENCE 508 AA; 56562 MW; 85823144F601FC6D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



OY 22 ITNNRLG 28  
 |||||  
 Db 373 ITNNRLG 379

```

RESULT 14
EDD_HELPJ STANDARD: PRT: 608 AA.
AC 082K83: 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoglucuronate dehydratase (EC 4.2.1.12) (6-phosphoglucuronate
  dehydratase).
GN EDD OR HPI1026.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria: Proteobacteria; epsilon subdivision; Helicobacter group;
  Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tumano P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
  gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- CATALYTIC ACTIVITY: 6-phospho-D-glucuronate = 2-dehydro-3-deoxy-6-
  phospho-D-glucuronate + H(2)O.
CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001530; AAC06597.1; -
DR InterPro; IPR000581; ILVD_EDD.
DR Pfam; PF00920; ILVD_EDD; 1.
DR ProDom; PD002691; ILVD_EDD; 1.
DR PROSITE; PS00886; ILVD_EDD_1; 1.
DR PROSITE; PS00887; ILVD_EDD_2; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 608 AA; 66603 MW; 978A046F3AE15F98 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 608;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 LIDEXSI 44
  |||||
Db 276 LIDEXSI 282

RESULT 15
EDD_HELPJ STANDARD: PRT: 608 AA.
AC P56111: 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoglucuronate dehydratase (EC 4.2.1.12) (6-phosphoglucuronate
  dehydratase).
GN EDD OR HPI100.
OS Helicobacter pylori (Campylobacter pylori).
  
```

```

OC Bacteria: Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glöck A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
  pylori."
RL Nature 388:539-547(1997).
CC -1- CATALYTIC ACTIVITY: 6-phospho-D-glucuronate = 2-dehydro-3-deoxy-6-
  phospho-D-glucuronate + H(2)O.
CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE00616; AAC08143.1; -
DR TIGR; HPI100; -
DR InterPro; IPR000581; ILVD_EDD.
DR Pfam; PF00920; ILVD_EDD; 1.
DR ProDom; PD002691; ILVD_EDD; 1.
DR PROSITE; PS00886; ILVD_EDD_1; 1.
DR PROSITE; PS00887; ILVD_EDD_2; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 608 AA; 66655 MW; 47EF7E62E3371F59 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 608;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 LIDEXSI 44
  |||||
Db 276 LIDEXSI 282
  
```

Search completed: August 15, 2002, 11:24:39  
 Job time: 686 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:27 ; Search time 96.53 Seconds

(without alignments)  
164.545 Million cell updates/sec

Title: US-08-981-087a-4

Perfect score: 143

Sequence: 1 NITSNRLTYGVGVIIIRKNG.....TSNGCWFMSFKSHGMOEN 143

Scoring table:

Gapop 60.0, Gapext 60.0

Word size: 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

1: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	143	100.0	143	18	AAW09017
2	143	100.0	431	18	AAW09014
3	143	100.0	432	22	ABO4096
4	143	100.0	432	22	ABO4096
5	143	100.0	645	22	AAE07894
6	143	100.0	685	22	AAE07893
7	143	100.0	862	22	AAE07890
8	143	100.0	887	22	AAE07892
9	143	100.0	1032	22	AAE07901
10	143	100.0	1059	21	AAV93309
11	143	100.0	1084	21	AAV93312

12	143	100.0	1092	22	AAE07900
13	92	64.3	432	21	AAV77138
14	24	16.8	448	19	AAW68398
15	8	5.6	449	22	AAW04095
16	8	5.6	449	21	AAV77137
17	8	5.6	449	21	AAW04094
18	8	5.6	452	19	AAW68396
19	7	4.9	96	21	AAW69310
20	7	4.9	104	22	AAW06221
21	7	4.9	114	22	AAE10807
22	7	4.9	245	20	AAV20027
23	7	4.9	265	20	AAV20026
24	7	4.9	289	21	AAW82013
25	7	4.9	335	21	AAV95073
26	7	4.9	387	12	AAE12576
27	7	4.9	387	13	AAE23652
28	7	4.9	451	19	AAW68395
29	7	4.9	518	22	AAW63292
30	7	4.9	554	19	AAW46873
31	7	4.9	554	20	AAV24976
32	7	4.9	632	20	AAW93890
33	7	4.9	693	20	AAV42169
34	7	4.9	797	22	AAU33386
35	7	4.9	768	22	AAU35171
36	7	4.9	819	22	ABW67656
37	7	4.9	907	20	AAW93889
38	7	4.9	907	21	AAV90682
39	7	4.9	907	21	AAV90687
40	7	4.9	1004	22	ABW64051
41	7	4.9	1049	22	ABW65142
42	7	4.9	1171	22	ABW64387
43	7	4.9	1265	22	ABG00708
44	7	4.9	1284	22	ABW71869
45	7	4.9	2280	22	ABW61650

#### ALIGNMENTS

RESULT 1	AAW09017	standard; Protein; 143 AA.
ID	AAW09017	
XX	AAW09017	
AC	AAW09017	
XX	AAW09017	
DT	31-MAR-1997	(first entry)
XX	AAW09017	
DE	Immunogenic type F botulinum toxin polypeptide (aa1136-1278).	
XX	AAW09017	
KX	Botulinum toxin; neurotoxin; BOBT/F; immunogen; vaccine; botulism.	
XX	AAW09017	
OS	Clostridium botulinum type F strain Langeland.	
XX	AAW09017	
PN	W09641881-AL.	
XX	AAW09017	
PD	27-DEC-1996.	
XX	AAW09017	
PF	12-JUN-1996; 96WO-GB01409.	
XX	AAW09017	
PR	12-JUN-1995; 95GB-0011909.	
XX	AAW09017	
PA	(MICR-) MICROBIOLOGICAL RES AUTHORITY.	
XX	AAW09017	
PI	Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;	
DR	WPI; 1997-065467/06.	
XX	AAW09017	
PT	Immunogenic type F botulinum toxin polypeptide(s) - allows	
XX	AAW09017	
PS	recombinant vaccine proth.	
XX	AAW09017	
XX	Claim 5; Page 19; 37pp: English.	
CC	Novel polypeptides (AAW09014-17) respectively comprise amino acids	

C. botulinum C2 tr  
Synthetic botulin  
Clostridium botuli  
Botulinum toxin hea  
Synthetic botulin  
Botulinum toxin hea  
Clostridium botuli  
HIV-1 non-subtype  
Human polypeptide  
Xanthomonas campe  
B. burgdorferi ant  
S. epidermidis ope  
Candida albicans p  
Ad41 long fibre pr  
Tak short fibre pr  
Clostridium botuli  
Drosophila melanog  
Bacillus thuringie  
Bacillus thuringie  
Human HG38 protein  
Human IGR5 protein  
Enterococcus faeca  
Drosophila melanog  
Human G protein-co  
Human G protein-co  
Drosophila melanog  
Drosophila melanog  
Novel human diageno  
Drosophila melanog  
Drosophila melanog

CC 848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a  
 CC type F botulinum neurotoxin (BoNT/F). They lack the L chain and  
 CC HN epitopes necessary for metalloprotease activity and toxin  
 CC internalisation. They are free of botulinum toxin activity but can  
 CC induce protective immunity to a type F botulinum toxin, making them  
 CC useful for vaccine prodn. Recombinant polypeptides can be  
 CC produced in transformed host cells, esp. as fusion proteins, e.g.  
 CC with maltose binding protein to facilitate purification.

SQ Sequence 143 AA;

Query Match 100.0%; Score 143; DB 18; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-140;  
 Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NIFSTRLTYGVEVIRKNSGTDISTDNFVRKNDLAYINVDVDRVRYADISIAKPE 60  
 DB 1 nifstrltygvevirknsgtdistdnfvrkndlayinvdvdvryadyadisiskpe 60  
 QY 61 KIILIRTSNSNSLGIIVWDSIGNNCTMNFQNNNGCIGLGFHSNNLVASSWYYNNI 120  
 DB 61 KIILIRTSNSNSLGIIVWDSIGNNCTMNFQNNNGCIGLGFHSNNLVASSWYYNNI 120  
 QY 121 RKNSSNGCFWSPFSKHEGQEN 143  
 DB 121 rknssngcfwspfskhegqen 143

#### RESULT 2

AAM09014  
 ID AAM09014 standard; Protein; 431 AA.

AC AAM09014;  
 DT 31-MAR-1997 (first entry)

DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).

KW Botulinum toxin; neurotoxin; BoBt/F; Immunogen; vaccine; botulism.

OS Clostridium botulinum type F strain Langeland.

PN WO9641881-A1.

PD 27-DEC-1996.

PF 12-JUN-1996; 96WO-GB01409.

PR 12-JUN-1995; 95GB-0011909.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;

DR WPI; 1997-065467/06.

DR N-PSDB; AAT48100.

PT Immunogenic type F botulinum toxin polypeptide(s) - allows

PS recombinant vaccine prodn.

PS Claim 5; Page 16-17; 37pp; English.

CC A polypeptide (AAM09014) comprises the heavy chain (amino acids  
 CC 848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be  
 CC produced using a synthetic gene (AAT48101) based on the natural  
 CC gene sequence (AAT48100) for the heavy chain. The polypeptides and  
 CC its fragments (see also AAM09015-17) lack the light chain and HN  
 CC epitopes necessary for metalloprotease activity and toxin  
 CC internalisation. They are free of botulinum toxin activity but can  
 CC induce protective immunity to a type F botulinum toxin, making them  
 CC useful for vaccine prodn. Recombinant polypeptides can be  
 CC produced in transformed host cells, esp. as fusion proteins, e.g.

CC with maltose binding protein to facilitate purification.

SQ Sequence 431 AA;

Query Match 100.0%; Score 143; DB 18; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-139;  
 Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NIFSTRLTYGVEVIRKNSGTDISTDNFVRKNDLAYINVDVDRVRYADISIAKPE 60  
 DB 289 nifstrltygvevirknsgtdistdnfvrkndlayinvdvdvryadyadisiskpe 348  
 QY 61 KIILIRTSNSNSLGIIVWDSIGNNCTMNFQNNNGCIGLGFHSNNLVASSWYYNNI 120  
 DB 349 KIILIRTSNSNSLGIIVWDSIGNNCTMNFQNNNGCIGLGFHSNNLVASSWYYNNI 408  
 QY 121 RKNSSNGCFWSPFSKHEGQEN 143  
 DB 409 rknssngcfwspfskhegqen 431

#### RESULT 3

AAB04096  
 ID AAB04096 standard; Protein; 432 AA.

AC AAB04096;

DT 11-APR-2001 (first entry)

DE Botulism toxin heavy chain C-terminal sequence (serotype F).

KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;

KW recombinant vector; antigen; immune response; vaccine; bacterium;

OS Synthetic.

OS Clostridium botulinum.

PN WO200067700-A2.

PD 16-NOV-2000.

PF 12-MAY-2000; 2000WO-US12890.

PR 12-MAY-1999; 99US-0133865.

PR 12-MAY-1999; 99US-0133866.

PR 12-MAY-1999; 99US-0133867.

PR 12-MAY-1999; 99US-0133868.

PR 12-MAY-1999; 99US-0133869.

PR 12-MAY-1999; 99US-0133873.

PR 29-JUL-1999; 99US-0146192.

PA (USDA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

DR WPI; 2001-016048/02.

DR N-PSDB; AAA54490.

PT New nucleic acids encoding the carboxy- or amino-terminal portions of

PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as

PT vaccine against botulism

PS Claim 3; Fig 9b; 73pp; English.

CC Botulism neurotoxins are translated as a single 150 kDa polypeptide

CC chain and then posttranslationally nicked, forming a di-chain

CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which

CC remain linked by a disulfide bond. Nucleic acids encoding the

CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy

CC chain of botulinum neurotoxin (BoNT) can be used in recombinant

CC expression vectors and expressed in transformed cells to produce

CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.

SQ Sequence 432 AA;

Query Match 100.0%; Score 143; DB 22; Length 432;  
Best Local Similarity 100.0%; Pred. No. 1.5e-139;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NIFSNRLTYGVEVIRKNGSDISNTDNFVRKNDLAYINVVDREYRLYADISIAKPE 60  
DB 290 nlfsnrltygvevirkngsdlsntdnfvrkndlayinvvdreyllyadislakpe 349

QY 61 KIIRLIRTSNNSNLGQIIYVDSIGNNCTMFONNNGNIGLGFHSNNLVASSWYYNNI 120  
DB 350 kikiirlrtsnnsnslgqiiyvdsignnctmfnngngnigllgfhsnnlvasswyyynni 409

QY 121 RKNTSSNGCFWFSISKHEGQEN 143  
DB 410 rkntssngcfwfsiskhegwen 432

RESULT 4  
AAB04103 standard; Protein: 432 AA.

XX AAB04103;  
XX 11-Apr-2001 (first entry)

DE Botulism toxin heavy chain C-terminal sequence (serotype F).

XX Botulism toxin; neurotoxin; heavy chain; recombinant expression;  
KW recombinant vector; antigen; immune response; vaccine; bacterium;  
XX infection.

OS Synthetic.  
OS Clostridium botulinum.

XX WO200067700-A2.  
XX 16-NOV-2000.  
XX 12-MAY-2000; 2000WO-US12890.  
XX 12-MAY-1999; 99US-0133865.  
XX 12-MAY-1999; 99US-0133866.  
XX 12-MAY-1999; 99US-0133867.  
XX 12-MAY-1999; 99US-0133868.  
XX 12-MAY-1999; 99US-0133869.  
XX 12-MAY-1999; 99US-0133873.  
XX 29-JUL-1999; 99US-0146192.

PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
XX N-PSDB; AAA54499.  
XX WPI: 2001-016048/02.  
XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
PT vaccine against botulism  
XX Disclosure; Fig 18b; 73pp; English.

XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked, forming a dichain which  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulism. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.

SQ Sequence 432 AA;

Query Match 100.0%; Score 143; DB 22; Length 432;  
Best Local Similarity 100.0%; Pred. No. 1.5e-139;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NIFSNRLTYGVEVIRKNGSDISNTDNFVRKNDLAYINVVDREYRLYADISIAKPE 60  
DB 290 nlfsnrltygvevirkngsdlsntdnfvrkndlayinvvdreyllyadislakpe 349

QY 61 KIIRLIRTSNNSNLGQIIYVDSIGNNCTMFONNNGNIGLGFHSNNLVASSWYYNNI 120  
DB 350 kikiirlrtsnnsnslgqiiyvdsignnctmfnngngnigllgfhsnnlvasswyyynni 409

QY 121 RKNTSSNGCFWFSISKHEGQEN 143  
DB 410 rkntssngcfwfsiskhegwen 432

RESULT 5  
AAE07894  
ID AAE07894 standard; Protein: 645 AA.

XX AAE07894;  
XX 01-NOV-2001 (first entry)

DE Modified clostridial heavy chain fragment #1.

XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
KW diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.

XX Chimeric - Corynebacterium diphtheriae.  
OS Chimeric - Clostridium botulinum.

XX WO200158936-A2.  
XX 16-AUG-2001.  
XX 04-DEC-2000; 2000WO-GB04644.  
XX 02-DEC-1999; 99GB-0028530.  
XX 07-APR-2000; 2000GB-0008658.  
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX Shone CC, Sutton JM, Silman N;  
XX WPI: 2001-514643/56.  
XX New non toxic polypeptide for delivery of a therapeutic agent for the  
PT treatment of a CNS disorder comprising a binding domain that  
PT translocates the therapeutic agent into the neuronal cells -

```
XX PS Example 2; Page 44; 50pp; English.
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain of botulinum neurotoxin type F
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 645 AA;
Query Match 100.0%; Score 143; DB 22; Length 645;
Best Local Similarity 100.0%; Pred. No. 2.2e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NIFSNTRLYGVGEVYIRKNGSTDISTDNFVRKNDLAYINVVDREVRYADISIAKPE 60
Db 503 nifsntrlygvgeviirkngstdistdnfvrkndlayinvvdrevryadiadisiakpe 562
QY 61 KIILKIRTSNNSNSLGOIYWDSIGNNCTMNFQNNNGSINGLGFHSNNLVASSWYYNNI 120
Db 563 kiiiklirtsnnsnslgoiiwdsignnctmfnqnnngsinglgfhsnnlvasswyyynni 622
QY 121 RKNTSSNGCFWSPFSKHEGMOEN 143
Db 623 rkntssngcfwspfskhegwgen 645
RESULT 6
AAE07893
ID AAE07893 standard; Protein; 685 AA.
AC AAE07893;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Influenza virus.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
```

```
XX PS New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
SQ Sequence 685 AA;
Query Match 100.0%; Score 143; DB 22; Length 685;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NIFSNTRLYGVGEVYIRKNGSTDISTDNFVRKNDLAYINVVDREVRYADISIAKPE 60
Db 543 nifsntrlygvgeviirkngstdistdnfvrkndlayinvvdrevryadiadisiakpe 602
QY 61 KIILKIRTSNNSNSLGOIYWDSIGNNCTMNFQNNNGSINGLGFHSNNLVASSWYYNNI 120
Db 603 kiiiklirtsnnsnslgoiiwdsignnctmfnqnnngsinglgfhsnnlvasswyyynni 662
QY 121 RKNTSSNGCFWSPFSKHEGMOEN 143
Db 663 rkntssngcfwspfskhegwgen 685
RESULT 7
AAE07890
ID AAE07890 standard; Protein; 862 AA.
AC AAE07890;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin;
KW botulinum neurotoxin type F; BoNT/F.
XX
PN WO200158936-A2.
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
```

```

PR 07-APR-2000; 2000GB-0008658.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX PA
XX PI Shone CC, Sutton JM, Silman N;
XX WPI; 2001-514643/56.
XX
XX New non toxic polypeptide for delivery of a therapeutic agent for the
XX treatment of a CNS disorder comprising a binding domain that
XX translocates the therapeutic agent into the neuronal cells -
XX
XX Example 9; Page 40; 50pp; English.
XX
XX The invention relates to a non toxic polypeptide, for delivery of a
XX therapeutic agent to a neuronal cell, which comprises a binding domain
XX (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
XX as Hc) that binds to the neuronal cell and a translocation domain (amino
XX terminal half of HC, designated as HN), that translocates the therapeutic
XX agent into the neuronal cell, where the translocation domain is not a HN
XX domain of a clostridial neurotoxin and is not a fragment or derivative of
XX a HN domain of a clostridial toxin. Polypeptides of the invention are
XX useful for the treatment of a disease state associated with neuronal
XX cells. The polypeptide constructs are useful for delivering therapeutic
XX substances to neuronal cells. They are useful to treat disorders of the
XX CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
XX and infection. They are also useful in gene therapy. The present sequence
XX is modified clostridial heavy chain-superoxide dismutase conjugate.
XX This conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
XX Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
XX translocation domain from diphtheria neurotoxin and a neuronal cell-
XX specific binding domain from botulinum neurotoxin type F (BoNT/F).
XX
XX Sequence 862 AA:
SQ
Query Match 100.0%; Score 143; DB 22; Length 862;
Best Local Similarity 100.0%; Pred. No. 2.9e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NIESENTRYGVETIRKNGSTDISNTDNFVRKNDLAYINVDVDEYRLYADISIAKPE 60
DB 720 nlfsvrlrylgvevllrkngstdlnctdnfvrkndlayinvdvdveyrlyadisiakpe 779
QY 61 KIKLIRTSNNSNSLGGIIVYDSTGNCTMNFONNNGNIGLGFHSNNLVASWYNNI 120
DB 780 kllklirtsnnsnslggilvmdsfgnctmfnngngniglgfhsnnlvasswyynni 839
QY 121 RKNSSNGCFWFSFKHGWOEN 143
DB 840 rknssngcfwfsfkhgwgwn 862

```

RESULT 8  
AAE07892  
ID AAE07892 standard; Protein; 887 AA.  
AC AAE07892;  
DT 01-NOV-2001 (first entry)  
XX Modified clostridial heavy chain-superoxide dismutase conjugate #4.  
XX  
XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
XX tumour; infection; neurodegenerative disease; gene therapy; chimeric;  
XX superoxide dismutase; SOD; diphtheria neurotoxin; human;  
XX botulinum neurotoxin type F; BoNT/F.  
XX  
XX Chimeric - Homo sapiens.  
XX Chimeric - Bacillus stearothermophilus.  
XX Chimeric - Corynebacterium diphtheriae.  
XX Chimeric - Clostridium botulinum.  
XX Chimeric - Synthetic.

```

XX W0200158936-A2.
XX PN
XX PD 16-AUG-2001.
XX PP
XX PF 04-DEC-2000; 2000WO-GB04644.
XX PR 02-DEC-1999; 99GB-0028530.
XX PR 07-APR-2000; 2000GB-0008658.
XX PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX PI Shone CC, Sutton JM, Silman N;
XX WPI; 2001-514643/56.
XX
XX New non toxic polypeptide for delivery of a therapeutic agent for the
XX treatment of a CNS disorder comprising a binding domain that
XX translocates the therapeutic agent into the neuronal cells -
XX
XX Example 9; Page 42; 50pp; English.
XX
XX The invention relates to a non toxic polypeptide, for delivery of a
XX therapeutic agent to a neuronal cell, which comprises a binding domain
XX (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
XX as Hc) that binds to the neuronal cell and a translocation domain (amino
XX terminal half of HC, designated as HN), that translocates the therapeutic
XX agent into the neuronal cell, where the translocation domain is not a HN
XX domain of a clostridial neurotoxin and is not a fragment or derivative of
XX a HN domain of a clostridial toxin. Polypeptides of the invention are
XX useful for the treatment of a disease state associated with neuronal
XX cells. The polypeptide constructs are useful for delivering therapeutic
XX substances to neuronal cells. They are useful to treat disorders of the
XX CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
XX and infection. They are also useful in gene therapy. The present sequence
XX is modified clostridial heavy chain-superoxide dismutase conjugate.
XX This conjugate comprises a mitochondrial leader sequence from human
XX Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
XX linker that can be cleaved by thrombin, translocation domain from
XX diphtheria neurotoxin and a neuronal cell-specific binding domain from
XX botulinum neurotoxin type F (BoNT/F).
XX
XX Sequence 887 AA:
SQ
Query Match 100.0%; Score 143; DB 22; Length 887;
Best Local Similarity 100.0%; Pred. No. 3e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NIESENTRYGVETIRKNGSTDISNTDNFVRKNDLAYINVDVDEYRLYADISIAKPE 60
DB 745 nlfsvrlrylgvevllrkngstdlnctdnfvrkndlayinvdvdveyrlyadisiakpe 804
QY 61 KIKLIRTSNNSNSLGGIIVYDSTGNCTMNFONNNGNIGLGFHSNNLVASWYNNI 120
DB 805 kllklirtsnnsnslggilvmdsfgnctmfnngngniglgfhsnnlvasswyynni 864
QY 121 RKNSSNGCFWFSFKHGWOEN 143
DB 865 rknssngcfwfsfkhgwgwn 887

```

RESULT 9  
AAE07901  
ID AAE07901 standard; Protein; 1032 AA.  
AC AAE07901;  
DT 01-NOV-2001 (first entry)  
XX C. botulinum C2 translocation domain with BoNT/F-binding domain #2.  
XX  
XX Neuronal cell; binding domain; translocation domain; stroke; epilepsy;

```
KW  tumour; infection; neurodegenerative disease; gene therapy;
KM  botulinum neurotoxin type F; BoNT/F.
XX
XX  Clostridium botulinum.
OS
XX  WO200158936-A2.
XX
XX  16-AUG-2001.
XX
XX  04-DEC-2000; 2000WO-GB04644.
XX
XX  02-DEC-1999; 99GB-0028530.
XX  07-APR-2000; 2000GB-0008658.
XX
XX  (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
XX  Shone CC, Sutton JM, Silman N;
XX
XX  WPI; 2001-514643/56.
XX
XX  New non toxic polypeptide for delivery of a therapeutic agent for the
XX  treatment of a CNS disorder comprising a binding domain that
XX  translocates the therapeutic agent into the neuronal cells -
XX
XX  Example 2; Page 48; 50pp; English.
XX
XX  The invention relates to a non toxic polypeptide, for delivery of a
XX  therapeutic agent to a neuronal cell, which comprises a binding domain
XX  (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
XX  as HC) that binds to the neuronal cell and a translocation domain (amino
XX  terminal half of HC, designated as HN), that translocates the therapeutic
XX  agent into the neuronal cell, where the translocation domain is not a HN
XX  domain of a clostridial neurotoxin and is not a fragment or derivative of
XX  a HN domain of a clostridial toxin. Polypeptides of the invention are
XX  useful for the treatment of a disease state associated with neuronal
XX  cells. The polypeptide constructs are useful for delivering therapeutic
XX  substances to neuronal cells. They are useful to treat disorders of the
XX  CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
XX  and infection. They are also useful in gene therapy. The present sequence
XX  is C. botulinum C2 enterotoxin translocation domain with botulinum
XX  neurotoxin type F (BoNT/F) binding domain used in the exemplification of
XX  the invention.
XX
XX  Sequence 1032 AA:
SQ
Query Match 100.0%; Score 143; DB 22; Length 1032;
Best Local Similarity 100.0%; Pred. No. 3.4e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NIFSNTRLTYGVEVYIRKNGSTDISNTDNFVRKNDLAYINVVDREVRLYADISIAPK 60
Db 890 nlfntnrltygvevylrkngstdisntdnfvrkndlayinvvdrevrlyadisiakpe 949
QY 61 KIILIRTSNNSNLSGQIIVWDSIGNNCTMNFQNNNGNIGLLGFHSNVLVASSWYNNI 120
Db 950 kiilirtsnsnns19qilivndsigntcmfngnngn19l1gfhsnvlvasswyyni 1009
QY 121 RKNTSSNCFWFSFSKEHGQEN 143
Db 1010 rktssngcfwfsfskshgwn 1032
RESULT 10
AAV93309
ID AAV93309 standard; protein; 1059 AA.
XX
XX  AAV93309;
XX
XX  04-SEP-2000 (first entry)
XX
XX  A manganese superoxide dismutase (Mn-SOD) construct.
XX
```

```
KW  Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KM  neuronal cell targeting component; NCIC; neuronal disease;
XX  oxidative stress; ischemic stroke; trauma; Parkinson's disease;
XX  Huntington's disease; motor neurone disease;
XX  botulinum neurotoxin serotype F.
XX
XX  Synthetic.
OS  Bacillus stearothermophilus.
XX  Clostridium botulinum.
XX
XX  WO200028041-A1.
XX
XX  18-MAY-2000.
XX
XX  05-NOV-1999; 99WO-GB03699.
XX
XX  05-NOV-1998; 98GB-0024282.
XX
XX  (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
XX  Shone CC, Sutton JM, Hallis B, Silman N;
XX
XX  WPI; 2000-376553/32.
XX
XX  Novel composition, comprising superoxide dismutase linked by a
XX  cleavable linker to a neuronal cell targeting component useful for
XX  delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
XX  Disclosure; Page 48-51; 65pp; English.
XX
XX  The present sequence represents a construct of the invention, comprising
XX  a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
XX  can be cleaved by thrombin, and a heavy chain derived from botulinum
XX  neurotoxin serotype F. The specification describes a composition for
XX  delivery of SOD to neuronal cells. The composition comprises SOD linked,
XX  by a cleavable linker, to a neuronal cell targeting component (NCIC).
XX  This component has a domain that binds to a neuronal cell and a
XX  domain that translocates the SOD of the composition into the neuronal
XX  cell. After translocation, the linker is cleaved to release the SOD.
XX  The composition is useful for treating neuronal diseases caused or
XX  augmented by oxidative stress, such as ischemic stroke, trauma,
XX  Parkinson's disease, Huntington's disease and motor neurone diseases.
XX
XX  Sequence 1059 AA:
SQ
Query Match 100.0%; Score 143; DB 21; Length 1059;
Best Local Similarity 100.0%; Pred. No. 3.5e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NIFSNTRLTYGVEVYIRKNGSTDISNTDNFVRKNDLAYINVVDREVRLYADISIAPK 60
Db 917 nlfntnrltygvevylrkngstdisntdnfvrkndlayinvvdrevrlyadisiakpe 976
QY 61 KIILIRTSNNSNLSGQIIVWDSIGNNCTMNFQNNNGNIGLLGFHSNVLVASSWYNNI 120
Db 977 kiilirtsnsnns19qilivndsigntcmfngnngn19l1gfhsnvlvasswyyni 1036
QY 121 RKNTSSNCFWFSFSKEHGQEN 143
Db 1037 rktssngcfwfsfskshgwn 1059
RESULT 11
AAV93312
ID AAV93312 standard; protein; 1084 AA.
XX
XX  AAV93312;
XX
XX  04-SEP-2000 (first entry)
XX
XX  A manganese superoxide dismutase (Mn-SOD) construct.
XX
```



KM Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
 KM neuronal cell targeting component; NCTC; neuronal disease;  
 KM oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
 KM Huntington's disease; motor neuron disease;  
 KM botulinum neurotoxin serotype F.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 OS Bacillus stearothermophilus.  
 OS Clostridium botulinum.  
 XX  
 PN WO200028041-A1.  
 XX  
 PD 18-MAY-2000.  
 XX  
 PF 05-NOV-1999; 99WO-GB03699.  
 XX  
 PR 05-NOV-1998; 98GB-0024282.  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 XX  
 PI Shone CC, Sutton JM, Hallis B, Silman N;  
 XX  
 DR WPI; 2000-376553/32.  
 XX  
 PT Novel composition, comprising superoxide dismutase linked by a  
 PT cleavable linker to a neuronal cell targeting component useful for  
 PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
 XX  
 PS Disclosure; Page 57-60; 65pp; English.  
 XX  
 CC The present sequence represents a construct of the invention, comprising  
 CC a mitochondrial leader sequence from human manganese superoxide  
 CC dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker  
 CC that can be cleaved by chymotrypsin, and a heavy chain derived from  
 CC botulinum neurotoxin serotype F. The specification describes a  
 CC composition for delivery of SOD to neuronal cells. The composition  
 CC comprises SOD linked, by a cleavable linker, to a neuronal cell  
 CC targeting component (NCTC). This component has a domain that binds  
 CC to a neuronal cell and a domain that translocates the SOD of the  
 CC composition into the neuronal cell. After translocation, the linker  
 CC is cleaved to release the SOD. The composition is useful for treating  
 CC neuronal diseases caused or augmented by oxidative stress, such as  
 CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and  
 CC motor neuron diseases.  
 CC  
 SQ Sequence 1084 AA;  
 XX  
 Query Match 100.0%; Score 143; DB 21; Length 1084;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-139;  
 Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NFESENRLTYGVEVIRKNGSTDSNTDNFVRKNDLAYINVDVDEYRLXADISTAKPE 60  
 Db 942 nlfesntlygvevllrkngstdsntdnfvrkndlayinvdvdeyrlxadisakpe 1001  
 QY 61 KIKILRTSNNSSNGQITIVMDSIGNNCTMNFQNNNGNIGLGFHSNNLVASSWYNNI 120  
 Db 1002 kikiilrtsnnsnsgqitlvmdsigncntmfnqnnngniglgfhsnnlvasswyynni 1061  
 QY 121 RKNNTSSNGCFMWSFISKEHGMOEN 143  
 Db 1062 rknntssngcfmwsfiskehgmgwen 1084  
 XX  
 RESULT 12  
 AAEO7900  
 ID AAEO7900 standard; Protein; 1092 AA.  
 AC AAEO7900;  
 XX  
 DT 01-NOV-2001 (first entry)

XX  
 DE C. botulinum C2 translocation domain with BONT/F-binding domain #1.  
 XX  
 KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;  
 KM tumour; infection; neurodegenerative disease; gene therapy;  
 KM botulinum neurotoxin type F; BONT/F.  
 XX  
 OS Clostridium botulinum.  
 OS  
 PN WO200158936-A2.  
 XX  
 PD 16-AUG-2001.  
 XX  
 PF 04-DEC-2000; 2000WO-GB04644.  
 XX  
 PR 02-DEC-1999; 99GB-0028530.  
 XX  
 PR 07-APR-2000; 2000GB-0008658.  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 XX  
 PI Shone CC, Sutton JM, Silman N;  
 XX  
 DR WPI; 2001-514643/56.  
 XX  
 PT New non toxic polypeptide for delivery of a therapeutic agent for the  
 PT treatment of a CNS disorder comprising a binding domain that  
 PT translocates the therapeutic agent into the neuronal cells -  
 XX  
 PS Example 2; Page 47; 50pp; English.  
 XX  
 CC The invention relates to a non toxic polypeptide, for delivery of a  
 CC therapeutic agent to a neuronal cell, which comprises a binding domain  
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated  
 CC as Hc) that binds to the neuronal cell and a translocation domain (amino  
 CC terminal half of HC, designated as HN), that translocates the therapeutic  
 CC agent into the neuronal cell, where the translocation domain is not a HN  
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of  
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are  
 CC useful for the treatment of a disease state associated with neuronal  
 CC cells. The polypeptide constructs are useful for delivering therapeutic  
 CC substances to neuronal cells. They are useful to treat disorders of the  
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours  
 CC and infection. They are also useful in gene therapy. The present sequence  
 CC is C. botulinum C2 enterotoxin translocation domain with botulinum  
 CC neurotoxin type F (BONT/F) binding domain used in the exemplification of  
 CC the invention.  
 CC  
 SQ Sequence 1092 AA;  
 XX  
 Query Match 100.0%; Score 143; DB 22; Length 1092;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-139;  
 Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NIFSNNRLTYGVEVIRKNGSTDSNTDNFVRKNDLAYINVDVDEYRLXADISTAKPE 60  
 Db 950 nifsnrltygvevllrkngstdsntdnfvrkndlayinvdvdeyrlxadisakpe 1009  
 QY 61 KIKILRTSNNSSNGQITIVMDSIGNNCTMNFQNNNGNIGLGFHSNNLVASSWYNNI 120  
 Db 1010 kikiilrtsnnsnsgqitlvmdsigncntmfnqnnngniglgfhsnnlvasswyynni 1069  
 QY 121 RKNNTSSNGCFMWSFISKEHGMOEN 143  
 Db 1070 rknntssngcfmwsfiskehgmgwen 1092  
 XX  
 RESULT 13  
 AA77138  
 ID AA77138 standard; Protein; 432 AA.  
 AC AA77138;  
 XX  
 DT

XX	08-MAY-2000	(first entry)
XX	Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.	
DE	Botulinum neurotoxin; heavy chain; BoNT; serotype F;	
KW	C-terminal fragment; Venezuelan equine encephalitis virus replicon;	
KW	VEE; botulinum vaccine; diagnosis; drug screening.	
XX		
OS	Clostridium botulinum.	
OS	Synthetic.	
PN	WO200002524-A2.	
PD	20-JAN-2000.	
XX		
XX	09-JUL-1999; 99WO-US15570.	
PE		
XX	10-JUL-1998; 98US-0092415.	
PR	12-MAY-1999; 99US-0133870.	
XX		
PA	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.	
PI		
PI	Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;	
DR	WPI; 2000-160827/14.	
DR	N-PSDB; AAZ87216.	
PT	Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum	
PT	toxin serotypes A-G, is used for inducing an immune response against	
PT	botulinum -	
XX		
PS	Claim 27; Page -; 54pp; English.	
XX		
CC	The invention relates to novel vaccines that induce a protective immune	
CC	response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F	
CC	and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant	
CC	DNA construct comprising a vector, and at least one nucleic acid	
CC	fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT	
CC	serotypes A-G, in preferred embodiments of the invention, the vector is	
CC	a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of	
CC	this vector results in the production of large amounts of a protein	
CC	encoded by a sequence cloned into the replicon. The constructs are used	
CC	to produce vaccines against botulism. The proteins can also be used as	
CC	diagnostic tools for the diagnosis of botulism. The transformed host	
CC	cells can be used to analyse the effectiveness of drugs and agents which	
CC	inhibit toxin effects. The vaccine currently used against botulism is	
CC	dangerous and expensive to produce, and contains formalin, which is very	
CC	painful for the recipient. Also, the vaccine is incomplete, in that only	
CC	5 of the 7 serotypes are represented in the formulation. The novel	
CC	vaccine of overcomes these problems, as it is easily purified, and	
CC	available in large quantities. It is also expressed in the lymph nodes	
CC	for a better immune response. Sequences AAY7134-Y77139 represent	
CC	synthetic BoNT Hc fragments used in the present invention. The DNA	
CC	encoding these sequences had been optimised for codon usage for	
CC	expression in yeast. Note: This sequence is not given in the	
CC	specification, but is decoded from the BoNTF Hc DNA sequence given on	
XX	pages 45-46.	
XX		
SQ	Sequence 432 AA;	

```

RESULT 14
AAW68399 ID AAW68399 standard; Protein: 448 AA.
XX AC AAW68399;
XX DF 07-DEC-1998 (first entry)
XX DE Clostridium botulinum type F toxin C fragment.
XX KM Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
XX botulism; BoTf.
OS Clostridium botulinum serotype F strain 202F (ATCC 23387).
XX Synthetic.
XX FH Key Location/Qualifiers
FH Peptide 1..21
FT /note= "N-terminal His tag"
XX WO9808540-A1.
PD 05-MAR-1998.
XX PF 28-AUG-1997; 97WO-US15394.
XX PR 28-AUG-1996; 96US-0704159.
XX PA (OPHI-) OPHIDIAN PHARM INC.
XX PI Thalley BS, Williams JA;
XX DR WPI: 1998-230234/20.
XX N-PBDB; AAV30593.
XX PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX Example 48; Page 364-365; 428pp; English.
PS This is the amino acid sequence of the histidine-tagged C fragment
CC of Clostridium botulinum (202F strain) type F neurotoxin, encoded
CC by a DNA sequence (see AAV30593) in plasmid pETHisB. This vector
CC can be used to express BotC soluble C fragment in Escherichia
CC coli host cells, with the recombinant C fragment being purified on
CC an affinity column. The invention relates to recombinant proteins
CC derived from C. botulinum toxins, especially type B and type E
CC toxins. Methods are provided which allow for the isolation of
CC soluble recombinant proteins free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of
CC vaccines and antitoxins that are useful in the treatment of humans
CC and animals at risk of intoxication with clostridial toxin.
XX SQ Sequence 448 AA;

Query Match 16.8%; Score 24; DB 19; Length 448;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 SLGGIIVWDSIGNCTMNFQNNG 97
Db 380 sLgqIIvmsDsignctmfnngnng 403

RESUL.T 15
AAB04095
ID AAB04095 standard; Protein: 419 AA.
XX
```

AC	AAB04095;
DT	11-APR-2001 (first entry)
DE	Botulinum toxin heavy chain C-terminal sequence (serotype E).
XX	
XX	Botulinism; toxin; neurotoxin; heavy chain; recombinant expression; KW recombinant vector; antigen; immune response; vaccine; bacterium; infection.
OS	Synthetic.
XX	Clostridium botulinum.
XX	MO200067700-A2.
PN	
PD	16-NOV-2000.
XX	
PE	12-MAY-2000; 2000WO-0512890.
XX	
PR	12-MAY-1999; 99US-0133865.
XX	12-MAY-1999; 99US-0133867.
PR	12-MAY-1999; 99US-0133867.
XX	12-MAY-1999; 99US-0133868.
PR	12-MAY-1999; 99US-0133869.
XX	12-MAY-1999; 99US-0133873.
PR	29-JUL-1999; 99US-0146192.
XX	
PA	(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
PI	Smith LA, Byrne MP, Middlebrook JT, Lapenotiere H;
DR	WPI; 2001-016048/02.
XX	N-PDB: AAA34489.
XX	
PT	New nucleic acids encoding the carboxy- or amino-terminal portions of the heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine against botulinism
PS	
XX	Disclosure: Fig 8; 73pp: English.
CC	Botulinum neurotoxins are translated as a single 150 kDa polypeptide chain and then posttranslationally nicked, forming a dichain consisting of a 100 kDa heavy chain and a 50 kDa light chain which remain linked by a disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT) can be used in recombinant expression vectors and expressed in transformed cells to produce peptide antigens useful for eliciting an immune response to give protective immunity against botulinum neurotoxin, which causes botulism. The nucleic acids are expressible in a recombinant organism such as Escherichia coli or Pichia pastoris. The use of recombinant nucleic acids are advantageous since it eliminates the need to culture large quantities of hazardous toxin-producing bacterium. Production yield from the genetically engineered product is also high and cost of production is lower. The nucleic acids can be derived from Clostridium botulinum serotypes A-G.
XX	
SO	Sequence 419 AA:
Query Match	5.6%; Score 8; DB 22; Length 419;
Best Local Similarity	100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 85 GNACTMNF 92	
db 361 gnnactmnf 368	

Search completed: August 15, 2002, 11:12:28  
Job time: 320 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:14:07 : Search time 47.36 Seconds  
(without alignments)  
290.135 Million cell updates/sec

Title: US-08-981-087a-4

Perfect score: 143  
Sequence: 1 NISFNRLYGVGVIIIRNK.....TSSNCGFWSFKSEHGMQEN 143

Scoring table: OLIGO  
Gapop 60.0, Gapept 60.0

Searched: 283138 seqs, 96089334 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: PIR\_71:\*

1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24	16.8	1274	2	140813	neurotoxin type F
2	16	11.2	1268	2	S33411	botulinum neurotoxin
3	8	5.6	1252	2	S21178	botulinum neurotoxin
4	7	4.9	67	2	A11594	hypothetical protein
5	7	4.9	242	2	T27999	hypothetical protein
6	7	4.9	259	2	AF1544	RNA polymerase sigma factor
7	7	4.9	259	2	AG1186	RNA polymerase sigma factor
8	7	4.9	264	2	G70195	pyridoxal kinase (EC 2.7.1.62)
9	7	4.9	280	2	T47572	pyridoxal kinase (EC 2.7.1.62)
10	7	4.9	347	2	T19989	hypothetical protein
11	7	4.9	375	2	A83636	hypothetical protein
12	7	4.9	380	2	H72374	probable phosphotransferase
13	7	4.9	387	1	ERADY1	41k filer protein
14	7	4.9	387	1	ERADY1	41k filer protein
15	7	4.9	396	2	G69794	hypothetical protein
16	7	4.9	407	2	A99223	sugar phosphate nucleotidyl transferase
17	7	4.9	515	2	A42140	box B-binding factor
18	7	4.9	516	2	A44494	CAMP-responsive element binding protein
19	7	4.9	523	2	S46720	NAM8 protein - yeast
20	7	4.9	543	2	T39345	probable metallothionein
21	7	4.9	781	2	E64222	DNA topoisomerase
22	7	4.9	907	2	JF0176	orphan G protein-coupled receptor
23	7	4.9	907	2	JG0193	G protein-coupled receptor
24	7	4.9	1132	2	S37206	phycochrome - moss
25	7	4.9	1139	1	E64234	cytochrome-c
26	7	4.9	1251	1	JH0256	botulinum neurotoxin
27	7	4.9	1303	2	C87519	hypothetical protein
28	7	4.9	1434	2	T22202	hypothetical protein
29	7	4.9	1802	2	H88444	protein C2656.12 [

30	6	4.2	91	2	H82370	conserved hypothet
31	6	4.2	105	2	JH0239	ferritin precurs
32	6	4.2	108	2	R43497	hypothetical prote
33	6	4.2	114	2	T09490	hypothetical prote
34	6	4.2	117	2	B86601	CTF41 hypothetical
35	6	4.2	117	2	H72022	hypothetical prote
36	6	4.2	119	2	H90531	preprotein translo
37	6	4.2	123	2	B69351	hypothetical prote
38	6	4.2	124	2	T38142	hypothetical prote
39	6	4.2	127	2	A11748	hypothetical prote
40	6	4.2	133	2	S57492	cytochrome-c oxida
41	6	4.2	133	2	S57491	cytochrome-c oxida
42	6	4.2	133	2	S57493	cytochrome-c oxida
43	6	4.2	136	2	D70361	transcription regu
44	6	4.2	140	2	JC5003	lysozyme [BC 3.2.1
45	6	4.2	146	2	T32375	hypothetical prote

#### ALIGNMENTS

RESULT 1  
140813  
neurotoxin type F - Clostridium botulinum  
C:Species: Clostridium botulinum  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999  
C:Accession: 140813; S48108  
R:East A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, F.E.M. Microbiol. Lett. 96, 225-230, 1992  
A:Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.  
A:Reference number: 140644  
A:Accession: 140813  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1274 <RES>  
A:Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867  
R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
A:Reference number: S48103; MUID:94013372  
A:Accession: S48108  
A>Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 634-1002 <CAM>  
A:Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 16.8%; Score 24; DB 2; Length 1274;  
Best Local Similarity 100.0%; Pred. No. 1.3e-16;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 SLGGIIIVMDSIGNNCTNFFONNG 97  
Db 1206 SLGGIIIVMDSIGNNCTNFFONNG 1229

RESULT 2  
S33411  
botulinum neurotoxin type F - Clostridium baratti  
C:Species: Clostridium baratti  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999  
C:Accession: S33411; S31860  
R:Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, F.E.M. Microbiol. Lett. 108, 175-182, 1993  
A:Title: Nucleotide sequence of the gene coding for Clostridium baratti type F neuroto  
A:Reference number: S33411; MUID:93352228  
A:Accession: S33411  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1268 <THO>  
A:Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139

C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 11.2%; Score 16; DB 2; Length 1268;  
Best Local Similarity 100.0%; Pred. No. 3; 1e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 127 NGCFWFSFISKEHGWOE 142

Db 1253 NGCFWFSFISKEHGWOE 1268

#### RESULT 3

S21178 botulinum neurotoxin type E precursor - Clostridium botulinum

C:Species: Clostridium botulinum  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Oct-1999  
C:Accession: S21178; S48107; JH0257; B35294; A60027; S18111

R:Wheeler, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.

Eur. J. Biochem. 204, 657-667, 1992

A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin  
A:Reference number: S21178; MUID:92174922

A:Accession: S21178

A:Molecule type: DNA

A:Residues: 1-1252 <MHE>

A:Cross-references: EMBL:X62683; NID:940397; PTDN:CAA44558.1; PID:940398

R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
A:Reference number: S48103; MUID:94013372

A:Accession: S48107

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 616-982 <CAW>

A:Cross-references: EMBL:X70815; NID:9407786; PTDN:CAA50146.1; PID:9407787

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.

Biochem. Biophys. Res. Commun. 183, 107-113, 1992

A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type  
A:Reference number: JH0256; MUID:92181428

A:Accession: JH0257

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-176, 'R', 178-197, 'C', 199-339, 'R', 341-772, 'T', 774-962, 'FE', 965-966, 'R', 968-1

A:Cross-references: EMBL:X62089; NID:940393; PTDN:CAA43999.1; PID:940394

A:Experimental source: strain Beluga

R:Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.

J. Biol. Chem. 265, 9153-9158, 1990

A:Title: The complete sequence of botulinum neurotoxin type A and comparison with other  
A:Reference number: A35294; MUID:90264400

A:Accession: B35294

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-176, 'R', 178-252 <BIN>

A:Experimental source: strain Beluga

R:Gimenez, J.A.; Dasgupta, B.R.

Biochimie 72, 213-217, 1990

A:Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the sh  
A:Reference number: A60027; MUID:90344918

A:Accession: A60027

A:Molecule type: protein

A:Residues: 420-427 <GIN>

A:Experimental source: strain Beluga

A:Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin

C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu

C:Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

F:1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>

F:423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>

F:412-426/Disulfide bonds: #status predicted

Query Match 5.6%; Score 8; DB 2; Length 1252;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 GNNCTMNF 92

Db 1194 GNNCTMNF 1201

#### RESULT 4

A11594 hypothetical protein lln1298 [imported] - Listeria innocua (strain Clp11262)

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: A11594

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baguerio, F.; Berche, P.; Bloec  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussureget, O.; Ertlan, K.D.; Fshh,  
D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maltournam, A.;  
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: A11594

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-67 <GLA>

A:Cross-references: GB:AL592022; PTDN:CAC96529.1; PID:g16413771; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lln1298

Query Match 4.9%; Score 7; DB 2; Length 67;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 INVVD RD 45

Db 31 INVVD RD 37

#### RESULT 5

T27999 hypothetical protein ZK795.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000

C:Accession: T27999

R:Percy, C.

submitted to the EMBL Data Library, December 1996

A:Reference number: Z20453

A:Accession: T27999

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-242 <NLI>

A:Cross-references: EMBL:283246; PTDN:CAB0542.1; GSPDB:GN00022; CESP:ZK795.1

A:Experimental source: clone ZK795

C:Genetics:

A:Gene: CESP:ZK795.1

A:Map position: 4

A:Inserts: 36/1; 59/3; 88/2; 116/1

C:Superfamily: Caenorhabditis elegans hypothetical protein ZK795.1

Query Match 4.9%; Score 7; DB 2; Length 242;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 IAKPEKI 62

Db 147 IAKPEKI 153

```

RESULT 6
AF1544
RNA polymerase sigma-37 factor (sigma-B) [imported] - Listeria innocua (strain Clp11262
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1544
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fshih, H.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehländ,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1544
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <GLA>
A:Cross-references: GB:AU592022; PIDN:CAC96126.1; PID:g1641344; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: sigB
C:Superfamily: transcription sigma factor G; transcription initiation factor sigma katF

Query Match 4.9%; Score 7; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 GNIGLIG 104
|||||
Db 63 GNIGLIG 69

RESULT 7
AG1186
RNA polymerase sigma-37 factor (sigma-B) [imported] - Listeria monocytogenes (strain EGD
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AG1186
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fshih, H.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehländ,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AG1186
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC98973.1; PID:g16410298; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: sigB
C:Superfamily: transcription sigma factor G; transcription initiation factor sigma katF

Query Match 4.9%; Score 7; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 GNIGLIG 104
|||||
Db 63 GNIGLIG 69

RESULT 8
G70195
pyridoxal kinase (pdxK) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (lyme disease spirochete)

```

```

C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70195
R:Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kellavag, A.R.; Quackenbush, J.; Salberg, S.; Hanson, M.; Vu
Boman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 380, 580-585, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:96065943
A:Accession: G70195
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-264 <KLE>
A:Cross-references: GB:AE001176; GB:AE000783; NID:g2688699; PIDN:AAC67112.1; PID:g268
A:Experimental source: strain B31

Query Match 4.9%; Score 7; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 EKIKLI 66
|||||
Db 90 EKIKLI 96

RESULT 9
T47572
Machado-Joseph disease MJD1a-like protein - Arabidopsis thaliana
N:Alternate names: protein F24B22.90
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47572
R:Bloeker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Queller, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223016
A:Accession: T47572
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <BL0>
A:Cross-references: EMBL:AL132957
A:Experimental source: cultivar Columbia; BAC clone F24B22
C:Genetics:
A:Map position: 3
A:Introns: 85/3
A:Note: F24B22.90

Query Match 4.9%; Score 7; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 IRKNSS 126
|||||
Db 204 IRKNSS 210

RESULT 10
T19989
hypothetical protein C47B2.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C:Accession: T19989
R:Kershaw, J.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z19208
A:Accession: T19989
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-347 <WTL>
A:Cross-references: EMBL:Z99709; PIDN:CAB1661.1; GSPDB:GN00019; CESP:C47B2.6
A:Experimental source: clone C47B2
C:Genetics:

```

A:Gene: CESP:C47B2.6  
A:Map position: 1  
A:Introns: 39/1; 115/3; 258/3; 296/3  
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology

Query Match 4.9%; Score 7; DB 2; Length 347;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 107 SNNLVS 113  
Db 104 SNNLVS 110

RESULT 11  
A83636  
hypothetical protein PA0063 [Imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83636  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Badian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.  
A:Reference number: A82950; MUID:20437337  
A:Accession: A83636  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-375 <STO>  
A:Cross-references: GB:AE004446; GB:AE004091; NID:g9945886; PIDN:AAG03453.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0063

Query Match 4.9%; Score 7; DB 2; Length 375;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 YRLYADI 54  
Db 107 YRLYADI 113

RESULT 12  
H72374  
probable phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) carbon dioxide-fixation  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72374  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Thermotoga maritima*.  
A:Reference number: A72200; MUID:99287316  
A:Accession: H72374  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-380 <ARN>  
A:Cross-references: GB:AE001723; GB:AE000512; NID:g4980953; PIDN:AAD35542.1; PID:g498096  
A:Experimental source: strain MS8  
C:Genetics:  
A:Gene: TM0447  
C:Superfamily: phosphoribosylaminoimidazole carboxylase carbon dioxide-fixation chain; F  
C:Keywords: carbon-carbon lyase; carboxy-lyase  
F:15-349/Domain: phosphoribosylaminoimidazole carboxylase carbon dioxide-fixation chain

Query Match 4.9%; Score 7; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 41 VVDRVE 47  
Db 349 VVDRVE 355

RESULT 13  
41K fiber protein - human adenovirus 41  
ERADN1  
A:Note: host Homo sapiens (man)  
C:Species: Mastadenovirus h41 (human adenovirus 41)  
C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999  
C:Accession: S09217; A45352; S20695  
R:Pieniazek, N.J.; Stenenda, S.B.; Pieniazek, D.; Velarde Jr., J.; Luftig, R.B.  
A>Title: Human enteric adenovirus type 41 (Tak) contains a second fiber protein gene.  
A:Reference number: S09217; MUID:90245595  
A:Accession: S09217  
A:Molecule type: DNA  
A:Residues: 1-387 <PIE>  
A:Cross-references: EMBL:X17016; NID:g58478; PIDN:CAA34882.1; PID:g58479  
A:Experimental source: strain Tak  
R:Kidd, A.H.; Erasmus, M.J.; Tiemessen, C.T.  
A>Title: Fiber sequence heterogeneity in subgroup F adenoviruses.  
A:Reference number: A45352; MUID:91021015  
A:Accession: A45352  
A:Molecule type: DNA  
A:Residues: 337-387 <KID>  
A:Cross-references: GB:M60327; NID:g209932; PIDN:AAA42505.1; PID:g209933  
A:Experimental source: strain FB585  
R:Pieniazek, N.J.; Stenenda, S.B.; Pieniazek, D.; Velarde Jr., J.; Luftig, R.B.  
A>Description: Characterisation of the early region E3 of the human enteric adenovirus 41.  
A:Reference number: S20688  
A:Accession: S20695  
A:Molecule type: DNA  
A:Residues: 1-58 <PI2>  
A:Cross-references: EMBL:X52198; NID:g58660; PIDN:CAA36450.1; PID:g58668  
C:Superfamily: adenovirus fiber protein  
C:Keywords: early protein

Query Match 4.9%; Score 7; DB 1; Length 387;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 NSNNSLG 76  
Db 89 NSNNSLG 95

RESULT 14  
ERADY4  
41K fiber protein - human adenovirus 40  
C:Species: Mastadenovirus h40 (human adenovirus 40)  
A:Note: host Homo sapiens (man)  
C>Date: 30-Jun-1990 #sequence\_revision 31-Mar-1992 #text\_change 03-May-1996  
C:Accession: A40048; A30336  
R:Kidd, A.H.  
A>Title: Sequence characterization of the adenovirus 40 fiber gene.  
A:Reference number: A30336; MUID:89370295  
A:Accession: A30336  
A:Molecule type: DNA  
A:Residues: 163-387 <KID2>



A:Cross-references: GB:M2822  
C:Superfamily: adenovirus fiber protein  
C:Keywords: early protein

Query Match 4.9%; Score 7; DB 1; Length 387;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 NSNNSLG 76  
|||||||  
DB 89 NSNNSLG 95

## RESULT 15

hypothetical protein yerrH - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C:Accession: G69794

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertier  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier  
Jech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelid  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033

A:Accession: G69794

A>Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1396 <KUN>

A:Cross-references: GB:Z99107; GB:AL09126; NID:g2632866; PIDN:CAM12483.1; PID:g2632977

A:Experimental source: strain 168

C:Genetics:

A:Gene: yerrH

C:Superfamily: Bacillus subtilis hypothetical protein yerrH

Query Match 4.9%; Score 7; DB 2; Length 396;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GSTDISN 26  
|||||||  
DB 266 GSTDISN 272

Search completed: August 15, 2002, 11:14:08  
Job time: 260 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:39 ; Search time 24.69 Seconds

(without alignments)  
224.237 Million cell updates/sec

Title: US-08-981-087a-4

Sequence: 1 NIESNRLTYGVEVIRRNQ.....TSSNCFMSISKHEGQEN 143

Scoring table: OLIGO

Gapop 60.0 , Gapect 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	16.8	1274	1 BXF_CLOBO	P30996 clostridium
2	7	4.9	387	1 FIR2_ADE40	P18048 human adeno
3	7	4.9	387	1 FIR2_ADE40	P16883 human adeno
4	7	4.9	515	1 CRBA_DROME	P29747 drosophila
5	7	4.9	523	1 NAME_YEAST	O00539 saccharomyc
6	7	4.9	781	1 PARC_MYCGE	P47446 mycoplasma
7	7	4.9	1132	1 PHY1_PHYPA	P36505 physcomitre
8	7	4.9	1139	1 HMM1_MYCGE	O49413 mycoplasma
9	7	4.9	1250	1 BXE_CLOBO	O00496 clostridium
10	7	4.9	1250	1 BXE_CLOBO	P30995 clostridium
11	6	4.2	64	1 Y05K_PPTA	P39238 bacteriopho
12	6	4.2	105	1 FBR_SACER	P24496 bacteriopho
13	6	4.2	109	1 COX1_SALIR	P29653 salmo trutt
14	6	4.2	117	1 Y884_CHLPP	O92722 chlamydia p
15	6	4.2	140	1 LYC_ANOGA	O17005 anopheles g
16	6	4.2	152	1 COX1_GEOSD	P29645 geophilus g
17	6	4.2	155	1 COX1_GOMVA	P29646 geophilus v
18	6	4.2	157	1 COX1_LEPSP	P29644 lepisosteus
19	6	4.2	157	1 COX1_SCAPL	P29654 scaphithyrc
20	6	4.2	158	1 COX1_POLSP	P29650 polyodon sp
21	6	4.2	160	1 COX1_MEGAT	P29648 megalops at
22	6	4.2	161	1 COX1_POMNI	P29652 pomoxis nig
23	6	4.2	163	1 COX1_LEPCC	O01423 lepisosteus
24	6	4.2	176	1 V19R_VACCV	O01223 vaccinia vi
25	6	4.2	178	1 RL6_HELPJ	O92486 helicobacte
26	6	4.2	178	1 RL6_HELPJ	P56034 helicobacte
27	6	4.2	184	1 COX1_PANBU	P29651 pandion bu
28	6	4.2	184	1 COX1_POLSX	P51243 polyporus
29	6	4.2	186	1 ATPD_PORPU	O90411 mesostigma
30	6	4.2	188	1 ATPD_PORPU	P29643 amia calva
31	6	4.2	192	1 COX1_AMICA	P70837 cochlidiobu
32	6	4.2	192	1 XN11_COCOA	O06562 cochlidiobu
33	6	4.2	221	1 XN11_COCOA	

34	5	4.2	245	1	EXPR_ERWCA	O47189 erwina car
35	5	4.2	257	1	NGF0_CHICK	P25433 gallus gall
36	5	4.2	273	1	YRF0_YEAST	P53177 saccharomyc
37	5	4.2	274	1	COX1_CHORI	P50668 choristoneu
38	5	4.2	274	1	COX1_CHORF	P50669 choristoneu
39	5	4.2	274	1	COX1_CHOC	P50670 choristoneu
40	5	4.2	274	1	COX1_CHORO	P50671 choristoneu
41	5	4.2	279	1	DMS0_HAEIN	P45002 haemophilus
42	5	4.2	279	1	LPXA_CHLPP	O92724 chlamydia p
43	5	4.2	279	1	NAD0_METTH	O27860 methanobact
44	5	4.2	280	1	LPXA_CHLMU	O93111 chlamydia m
45	5	4.2	280	1	LPXA_CHLIR	O84536 chlamydia t

#### ALIGNMENTS

RESULT 1

ID	BXF_CLOBO	STANDARD	PRT: 1274 AA.
AC	P30996;		
DT	01-JUL-1993 (Rel. 26, Created)		
DT	01-JUL-1993 (Rel. 26, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)		
DE	(Bontoxilysin F).		
GN	BoNT		
OS	Clostridium botulinum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OX	Clostridium.		
OX	NCBI_TaxID=1491;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 23387;		
RC	MEDLINE=93012902; PubMed=1398040;		
RA	East A.K., Richardson P.T., Allaway D., Collins M.D.,		
RA	Roberts T.A., Thompson D.E.;		
RT	*Sequence of the gene encoding type F neurotoxin of Clostridium		
RT	botulinum.";		
RT	FEMS Microbiol. Lett. 75:225-230(1992).		
RN	[2]		
RP	SEQUENCE OF 1-64 FROM N.A.		
RC	STRAIN=HOBBS FT10;		
RC	MEDLINE=94297488; PubMed=7764998;		
RA	East A.K., Collins M.D.;		
RT	*Conserved structure of genes encoding components of botulinum		
RT	neurotoxin complex M and the sequence of the gene coding for the		
RT	neurotoxin component in nonproteolytic Clostridium botulinum type F.";		
RT	Curr. Microbiol. 29:65-77(1994).		
RN	[3]		
RP	SEQUENCE OF 634-1002 FROM N.A.		
RC	MEDLINE=94013372; PubMed=8408542;		
RA	Campbell K., East A.K., Collins M.D.;		
RT	*Gene probes for identification of the botulinum neurotoxin gene and		
RT	specific identification of neurotoxin types B, E, and F.";		
RT	J. Clin. Microbiol. 31:2255-2262(1993).		
RN	[4]		
RP	IDENTIFICATION OF SUBSTRATE.		
RC	MEDLINE=94230352; PubMed=8175689;		
RA	Yamasaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,		
RA	Rogues B., Fyfe E.M., Suedhof T.C., Jahn R., Niemann H.;		
RT	*Cleavage of members of the synaptobrevin/VAMP family by types D and		
RT	F botulinum neurotoxins and tetanus toxin.";		
RT	J. Biol. Chem. 269:12764-12772(1994).		
CC	-I- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER		
CC	RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES. IS INTERNALIZED		
CC	AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD		
CC	WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT		
CC	INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC		
CC	ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-1-LYS-59		
CC	BOND OF SYNAPTOSOMAL VESICLE AND -2.		
CC	-I- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the		
CC	neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No		

```

CC detected action on small molecule substrates.
CC -1 SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M92906; AAA23263.1; -
DR EMBL: S73676; AAC60475.1; -
DR EMBL: X70820; CAA50151.1; -
DR EMBL: X70816; CAA50147.1; -
DR HSSP: P10845; 3BTA.
DR MEROPS: M27.002; -
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01742; Peptidase_M27.1.
DR PRINTS: PR00760; BONTOXILYSIN.
DR PRODOM: PD001963; BONTOXILYSIN.
DR PROSITE: PS00142; ZINC_PROTEASE.1.
DR Neurotoxin: Transmembrane; Hydrolase; Metalloprotease; Zinc.
KW CHAIN 1 436
FT CHAIN 1 436
FT METAL 437 1274
FT ACT_SITE 227 227
FT METAL 228 228
FT METAL 231 231
FT DISULFID 429 445
FT SEQUENCE 1274 AA; 146709 MW; 5B99756A7438B921 CRC64;

Query Match
Best Local Similarity 100.0%; Score 24; DB 1; Length 1274;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 SLGQIIYMSIGNNCTMNFONNG 97
DB 1206 SLGQIIYMSIGNNCTMNFONNG 1229

RESULT 2
FIB2_ADE40 STANDARD; PRT; 387 AA.
AC P18048;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Fiber protein 2.
OS Human adenovirus type 40.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28284;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DUGAN;
RX MEDLINE=94087748; PubMed=8263936;
RA Davidson A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT "The DNA sequence of adenovirus type 40.";
RL J. Mol. Biol. 234:1308-1316(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93297140; PubMed=8517033;
RA Kidd A.H., Chroboczek J., Cusack S., Ruigrok R.W.H.;
RT "Adenovirus type 40 virions contain two distinct fibers.";
RL Virology 192:73-84(1993).

```

```

RN [3]
RP SEQUENCE OF 167-387 FROM N.A.
RX MEDLINE=89370295; PubMed=2773314;
RA Kidd A.H., Erasmus M.J.;
RT "Sequence characterization of the adenovirus 40 fiber gene.";
RL Virology 172:134-144(1989).
CC -1 FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC -1 SUBUNIT: HOMOTRIMER (BY SIMILARITY).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: L19443; AAC13979.1; -
DR EMBL: M28822; AAA03233.1; -
DR PIR: A40048; ERADY4.
DR InterPro: IPR000939; Adeno_fiber2.
DR InterPro: IPR000978; Adeno_fiber_knob.
DR InterPro: IPR000931; Adeno_fibre.
DR Pfam: PF00541; adeno_fiber.1.
DR Pfam: PF00608; adeno_fiber2.5.
DR PRINTS: PR00307; ADENOVSFIBRE.
KW Fiber protein.
FT CONFLICT 226 226 G -> S (IN REF. 2 AND 3).
FT SEQUENCE 387 AA; 41346 MW; 11A3C1FED61A3ACB CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 387;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 NSNNSIG 76
DB 89 NSNNSIG 95

RESULT 3
FIB2_ADE41 STANDARD; PRT; 387 AA.
AC P16883;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Fiber protein 2.
OS Human adenovirus type 41.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10524;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TAK;
RX MEDLINE=90245595; PubMed=2336370;
RA Pieniazek N.J., Slemenda S.B., Pieniazek D., Velarde J. Jr.,
RA Luftig R.B.;
RT "Human enteric adenovirus type 41 (Tak) contains a second fiber
RT protein gene.";
RL Nucleic Acids Res. 18:1901-1901(1990).
RN [2]
RP SEQUENCE OF 337-387 FROM N.A.
RC STRAIN=FB585;
RX MEDLINE=91021015; PubMed=2219717;
RA Kidd A.H., Erasmus M.J., Tiemessen C.F.;
RT "Fiber sequence heterogeneity in subgroup F adenoviruses.";
RL Virology 179:139-150(1990).
CC -1 FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC -1 SUBUNIT: HOMOTRIMER (BY SIMILARITY).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as their content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
CC -----  
DR EMBL: X17016; CAA34882.1; -  
DR EMBL: M60327; AAA42505.1; -  
DR PIR: S09217; ERA0N1.  
DR PIR: A45352; A45352.  
DR HSSP: P11818; 1KNB.  
DR InterPro: IPR000936; Adeno\_fiber2.  
DR InterPro: IPR000931; Adeno\_fiber\_knob.  
DR InterPro: IPR000931; Adeno\_fibre.  
DR Pfam: PF00341; adeno\_fiber2\_1.  
DR Pfam: PF00608; adeno\_fiber2\_5.  
DR PRINTS: PR00307; ADENOVSFIBRE.  
KW Fiber protein.  
SQ SEQUENCE 387 AA; 41397 MW; 8652E785276573C7 CRC64;  
  
Query Match 4.9%; Score 7; DB 1; Length 387;  
Best Local Similarity 100.0%; Pred. No. 8.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 70 NSNNSLG 76  
| | | | | | | |  
DB 89 NSNNSLG 95  
  
RESULT 4  
CRAA\_DROME STANDARD; PRT; 515 AA.  
ID CRAA\_DROME  
AC P29747  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cyclic-AMP response element binding protein A (Box B binding factor-2)  
DE (BBF-2).  
GN CREBA OR BBF2.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON-R;  
RX MEDLINE=92192458; PubMed=1532159;  
RA Abel T., Bhatt R., Maniatis T.;  
RT "A Drosophila CREB/ATF transcriptional activator binds to both fat  
RT body- and liver-specific regulatory elements.";  
RL Gene Dev. 6:466-480(1992).  
CC  
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS TO FAT BODY-SPECIFIC  
CC ENHANCERS OF ALCOHOL DEHYDROGENASE (ADH) AND YOLK PROTEIN GENES.  
CC BBF-2 MAY PLAY A ROLE IN FAT BODY GENE EXPRESSION. IT BINDS THE  
CC CONSENSUS SEQUENCE 5'T(A/C)NACGTAA(T/G)C'-3'.  
CC -1- SUBUNIT: MAY BIND DNA AS HETERODIMERS WITH OTHER BZIP PROTEINS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: IN ALL CELL TYPES EXAMINED.  
CC -1- DEVELOPMENTAL STAGE: PRESENT THROUGHOUT DEVELOPMENT.  
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.  
CC  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
CC -----  
DR EMBL: X64429; CAA45771.1; -

DR PIR: S24542; S24542.  
DR PIR: A42140; A42140.  
DR TRANSFAC: T01603; -  
DR FlyBase: FBgn0004396; Creba.  
DR InterPro: IPR001871; bZIP.  
DR Pfam: PF00170; bZIP\_1.  
DR SMART: SM00338; BRIZ; 1.  
DR PROSITE: PS00036; bZIP\_BASIC; 1.  
KW Transcription regulation; Activator; DNA-binding; Nuclear protein.  
FT DNA\_BIND 448  
FT BASIC\_MOTIF 448  
FT LEOCTNE\_ZIPPER 503  
SQ SEQUENCE 515 AA; 56528 MW; 0E08FB9655200223 CRC64;  
  
Query Match 4.9%; Score 7; DB 1; Length 515;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 69 SNSNSNL 75  
| | | | | | | |  
DB 264 SNSNSNL 270  
  
RESULT 5  
NM88\_YEAST STANDARD; PRT; 523 AA.  
ID NM88\_YEAST  
AC O00539;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE NM88 protein.  
DE NM88 OR WRE2 OR YHR086W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R23/50;  
RX MEDLINE=92293106; PubMed=1603056;  
RA Ekwall K., Kermorgant M., Dujardin G., Groudinsky O.,  
RA Slonimski P.P.;  
RT "The NM88 gene in Saccharomyces cerevisiae encodes a protein with  
RT putative RNA binding motifs and acts as a suppressor of mitochondrial  
RT splicing deficiencies when overexpressed.";  
RL Mol. Gen. Genet. 233:136-144(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Isem S.-H., Hayashi A., Ajimura M., Ogawa H.;  
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
RA Du Z., Favali A., Fulton L., Gatlund S., Gelsel C., Kirsten J.,  
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
RA Latreille P., Louis E.J., Macri C., Maris E., Meneses S., Mouser L.,  
RA Nman M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,  
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
RA Vaudin M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
RT VIII.";  
RL Science 265:2077-2082(1994).  
CC  
CC -1- FUNCTION: ACTS AS A SUPPRESSOR OF MITOCHONDRIAL SPLICING  
CC DEFICIENCIES WHEN OVEREXPRESSED. COULD BE A NON-ESSENTIAL  
CC COMPONENT OF THE MITOCHONDRIAL SPLICING MACHINERY.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).  
CC  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64763; CAA46011.1; -
DR EMBL: D11461; BAA02016.1; -
DR EMBL: U00060; AAB68928.1; -
DR PIR: S22439; S22439.
DR PIR: S46720; S46720.
DR SGD: S0001128; NAM8.
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; rrm; 3.
DR SMART: SM00360; RRM; 3.
DR PROSITE: PS0102; RRM; 3.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW Nuclear protein; RNA-binding; Mitochondrion; mRNA processing;
Repeat.
FT DOMAIN 54 145 RNA-BINDING (RRM) 1.
FT DOMAIN 163 242 RNA-BINDING (RRM) 2.
FT DOMAIN 313 385 RNA-BINDING (RRM) 3.
FT CONFLICT 180 180 F -> L (IN REF. 1).
FT CONFLICT 208 209 GF -> VL (IN REF. 1).
SQ SEQUENCE 523 AA; 56972 MW; 64F198EEFB32A909 CRC64;

Query Match
Best Local Similarity 4.9%; Score 7; DB 1; Length 523;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 SNSNSL 75
Db 148 SNSNSL 154

RESULT 6
PARC MYCSE STANDARD; PRT; 781 AA.
AC P47446; Q49377.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase IV subunit A (EC 5.99.1.-).
GN PARC OR MG204.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE-96026346; PubMed-7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhlman J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tombl J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lueder T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1-479 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RA Bailey C.C., Younkins R., Huang W.M., Bott K.F.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBD databases.
CC -1- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME
CC SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.
CC PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION
CC OF A CIRCULAR DNA MOLECULE (BY SIMILARITY).
CC -1- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PAR.
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39700; AAC71422.1; -
DR EMBL: U25549; AAC43991.1; -
DR HSSP: P09097; LAB4.
DR TIGR: MG204; -
DR InterPro: IPR002205; DNA_topoisomIV.
DR Pfam: PF00521; DNA_topoisomIV; 1.
DR SMART: SM00434; TOP4c; 1.
KW Topoisomerase; Isomerase; DNA-binding; Complete proteome.
FT ACT SITE 122 122 P -> R (IN REF. 2).
FT CONFLICT 261 261
SQ SEQUENCE 781 AA; 88512 MW; F14319CEB305B437 CRC64;

Query Match
Best Local Similarity 4.9%; Score 7; DB 1; Length 781;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 KIIRLIR 67
Db 388 KIIRLIR 394

RESULT 7
PHY1_PHYPA STANDARD; PRT; 1132 AA.
ID PHY1_PHYPA
AC P36505;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytochrome 1.
GN PHY1.
OS Physcomitrella patens (Moss).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX NCBI_TaxID=3218;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94039823; PubMed-8224238;
RA Kolukisaoglu H.U., Brun B., Martin W.F., Schneider-Poetsch H.A.W.;
RT "Mosses do express conventional, distantly B-type-related
RT phytochromes. Phytochrome of Physcomitrella patens (Hedw.).";
RL FEBS Lett. 334:95-100(1993).
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
CC PFR INDICES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RUBULOSE-
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC PROTOCHLOROPHYLLIDE REDUCTASE, RRNA, ETC. IT ALSO CONTROLS
CC THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC -1- SUBUNIT: HOMODIMER.
CC -1- PFM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARR-T-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X75025; CAA52933.1; -.  
 DR PIR: S37206; S37206.  
 DR InterPro: IPR003018; GAF.  
 DR InterPro: IPR003594; HATPase\_C.  
 DR InterPro: IPR004359; HIS\_KIN\_sig.  
 DR InterPro: IPR003661; HIS\_KIN.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR InterPro: IPR001294; Phytochrome.  
 DR Pfam: PF01590; GAF; 1.  
 DR Pfam: PF02518; HATPase\_C; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR Pfam: PF00360; Phytochrome; 1.  
 DR Pfam: PF00512; Signal; 1.  
 DR PRINTS: PR01033; PHYTOCHROME.  
 DR SMART: SM00065; GAF; 1.  
 DR SMART: SM00387; HATPase\_C; 1.  
 DR SMART: SM00388; HSKA; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS50109; HIS\_KIN; 1.  
 DR PROSITE: PS50112; PAS; 2.  
 DR PROSITE: PS00245; PHYTOCHROME\_1; 1.  
 DR PROSITE: PS50046; PHYTOCHROME\_2; 1.  
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;  
 Repeat.  
 FT DOMAIN 610 681 PAS 1.  
 FT DOMAIN 744 815 PAS 2.  
 FT DOMAIN 895 1115 HISTIDINE KINASE.  
 FT BINDING 321 321 CHROMOPHORE (BY SIMILARITY).  
 SQ SEQUENCE 1132 AA; 125230 MW; ELDAD4D6DC9CDD16 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 1132;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 GGNIGL 103  
 Db 214 GGNIGL 220

RESULT 8  
 HMW1\_MYGE STANDARD; PRT; 1139 AA.  
 AC 049413; 049365;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome high molecular weight protein 1 (Cytochrome accessory  
 protein 1).  
 GN HMW1 OR MG312.  
 OS Mycoplasma genitalium.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 OC Mycoplasmataceae; Mycoplasma.  
 NCBI\_TaxID=2097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33530 / G-37;  
 RX MEDLINE=96026346; PubMed=7569993;  
 RA Fraser C.H., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
 RA Fleischmann R.D., Bult C.J., Kettlewell A.R., Sutton G., Kelley J.M.,  
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
 RA Nguyen D.T., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
 RA Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Luster T.S.,  
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
 RT "The minimal gene complement of Mycoplasma genitalium";  
 RL Science 270:397-403(1995).  
 RN [2]  
 RP SEQUENCE OF 721-847 FROM N.A.

RC STRAIN-ATCC 33530 / G-37;  
 RX MEDLINE=96075230; PubMed=8253680;  
 RA Peterson S.N., Hu P.-C., Boff K.F., Hutchison C.A. III;  
 RT "A survey of the Mycoplasma genitalium genome by using random  
 RT sequencing";  
 RL J. Bacteriol. 175:7918-7930(1993).  
 CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH  
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS  
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW  
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS  
 CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT  
 CC MEMBRANE (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U39712; AAC71534.1; -.  
 DR EMBL: U02261; MAD12527.1; -.  
 DR TIGR: MG312; -.  
 KW Cytochrome; Structural protein; Complete proteome.  
 SQ SEQUENCE 1139 AA; 130531 MW; 0011D3288C3D856 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 1139;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 SNSNSL 75  
 Db 450 SNSNSL 456

RESULT 9  
 BXL\_CLOBO STANDARD; PRT; 1250 AA.  
 AC 000496;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)  
 DE (Bontoxilysin E).  
 OS Clostridium botulinum.  
 OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BELUGA;  
 RX MEDLINE=92181428; PubMed=1543481;  
 RA Boullet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;  
 RT "Sequences of the botulinum neurotoxin E derived from Clostridium  
 RT botulinum type E (strain Beluga) and Clostridium butyricum (strains  
 RT ATCC 43181 and ATCC 43755).";  
 RL Biochem. Biophys. Res. Commun. 183:107-113(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92174922; PubMed=1541280;  
 RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;  
 RT "The complete amino acid sequence of the Clostridium botulinum type-E  
 RT neurotoxin, derived by nucleotide-sequence analysis of the encoding  
 RT gene";  
 RL Eur. J. Biochem. 204:657-667(1992).  
 RN [3]  
 RP SEQUENCE OF 1-251 FROM N.A.  
 RX MEDLINE=90264400; PubMed=2160960;  
 RA Binz T., Kurazono H., Wille M., Frevet J., Wernars K., Niemann H.;

RT "The complete sequence of botulinum neurotoxin type A and comparison  
 RT with other clostridial neurotoxins.";  
 RL J. Biol. Chem. 265:9153-9158(1990).  
 RN [4]  
 RP SEQUENCE OF 1-13.  
 RX MEDLINE=85197963; PubMed=3888113;  
 RA Schmidt J.J., Sathymoorthy V., Dasgupta B.R.;  
 RT Partial amino acid sequences of botulinum neurotoxins types B and  
 RT E.";  
 RL Arch. Biochem. Biophys. 238:544-548(1985).  
 RN [5]  
 RP SEQUENCE OF 419-426.  
 RX MEDLINE=90344918; PubMed=2116911;  
 RA Gimenez J.A., Dasgupta B.R.;  
 RT Botulinum neurotoxin type E fragmented with endoproteinase Lys-C  
 RT reveals the site trypsin nicks and homology with tetanus  
 RT neurotoxin.";  
 RL Biochimie 72:213-217(1990).  
 RN [6]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94063091; PubMed=8243676;  
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
 RA Benfenati F., Wilson M.C., Montecucco C.;  
 RT Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 RT COOH-terminal peptide bonds.";  
 RL FEBS Lett. 335:99-103(1993).  
 RN [7]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94124495; PubMed=8294407;  
 RA Binz T., Blaszi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 RA Jahn R., Niemann H.;  
 RT Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";  
 RL J. Biol. Chem. 269:1617-1620(1994).  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-1-ILE-  
 CC 181 BOND IN SNAP-25.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X62089; CAA43999.1; -;  
 DR EMBL: X62683; CAA44358.1; -;  
 DR PIR: A60027; A60027.  
 DR PIR: B35294; B35294.  
 DR PIR: JH0257; JH0257.  
 DR PIR: S08575; S08575.  
 DR PIR: S18111; S18111.  
 DR PIR: S21178; S21178.  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -;  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.

DR PRINTS: PR00760; BONTOXILYSIN.  
 DR PRODOM: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; zinc.  
 FT INIT.MET 0 FT  
 FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.  
 FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.  
 FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY);  
 FT ACT\_SITE 212 212 BY SIMILARITY.  
 FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 411 425 INTERCHAIN (PROBABLE).  
 FT CONFLICT 197 197 R -> G (IN REF. 2).  
 FT CONFLICT 339 339 R -> S (IN REF. 2 AND 3).  
 FT CONFLICT 772 772 I -> L (IN REF. 2).  
 FT CONFLICT 962 963 FE -> LQ (IN REF. 2).  
 FT CONFLICT 966 966 R -> A (IN REF. 2).  
 FT CONFLICT 1194 1194 N -> NN (IN REF. 2).  
 SQ SEQUENCE 1250 AA; 143712 MW; D9PC26DDA041EB4 CRC64;  
 Query Match 4.9%; Score 7; DB 1; Length 1250;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 99 NIGLGF 105  
 Db 1206 NIGLGF 1212  
 |||||  
 RESULT 10  
 BXE\_CLOBU BXE\_CLOBU STANDARD; PRT; 1250 AA.  
 ID P30995;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)  
 DE (Bontoxilysin E).  
 OS Clostridium butyricum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1492;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 43181, AND ATCC 43755;  
 RX MEDLINE=92181428; PubMed=1543481;  
 RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;  
 RT "Sequences of the botulinum neurotoxin E derived from Clostridium  
 RT botulinum type E (strain Beluga) and Clostridium butyricum (strains  
 RT ATCC 43181 and ATCC 43755).";  
 RL Biochem. Biophys. Res. Commun. 183:107-113(1992).  
 RP SEQUENCE OF 1-251 FROM N.A.  
 RC STRAIN=Bl6340;  
 RX MEDLINE=91237316; PubMed=2033376;  
 RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,  
 RA Yokosawa N., Yashiki T., Oguma K.;  
 RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum  
 RT type E toxin gene from Clostridium butyricum strain Bl6340.";  
 RL J. Gen. Microbiol. 137:519-525(1991).  
 RN [3]  
 RP SEQUENCE OF 1-48.  
 RC STRAIN=5262;  
 RA Gimenez J., Foley J., Dasgupta B.R.;  
 RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;  
 RT partial sequence and comparison.";  
 RL FASEB J. 2:A1750-A1750(1988).  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC



```

CC ENDOPEPTIDASE.
CC -1- CATALYTIC ACTIVITY: limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X62088; CAA43998.1; -.
CC EMBL: X53180; CAA37321.1; -.
CC PIR: JH0256; JH0256.
CC PIR: S16145; S16145.
CC HSSP: P10845; 3BTA.
CC MEROPS: M27.002; -.
CC InterPro: IPR000395; Bontoxilysin.
CC InterPro: IPR000130; Zn_mtpetidase.
CC Pfam: PF01742; Peptidase_M27_1.
CC PRINTS: PR00760; BONTOTOXILYSIN.
CC ProDom: PD001963; BONTOTOXILYSIN.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
CC FT INIT_MET 0
CC FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
CC FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
CC FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACET_SITE 212 212 BY SIMILARITY.
CC FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DISULFID 411 425 INTERCHAIN (PROBABLE).
CC FT CONFLICT 229 229 K -> M (IN REF. 2).
CC FT SEQUENCE 1250 AA; 143265 MW; 8171B5B2C312857 CRC64;

Query Match 4.9%; Score 7; DB 1; Length 1250;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 NIGLIGF 105
DB 1206 NIGLIGF 1212

RESULT 11
Y05K_BP74 STANDARD; PRT; 64 AA.
ID Y05K_BP74
AC P39238;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 7.6 kDa protein in mobd-r1 intergenic region.
GN Y05K OR MOB.D.3 OR TK.-7.
OS Bacteriophage T4.
OC Viruses: dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OC NCBI_TaxID=10665;
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RA Mzhavits N., Marusch E., Djavakhishvili T., Neitzel J., Peterson S.,
RA Awaja M., Eldemiller J., Canada D., Tracy J., Galibereath K.,
RA Paddison F., Anderson B., Stidham T., Blattner F., Kuter E.M.;
RA The 10.7 kb nonessential region of bacteriophage T4 between the

```

```

RT genes tk and nrdC: twenty new t4 genes, generally conserved among
RT T-seven phages."
RT Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kuter E., Aisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis".
RT Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76612; AAB26966.1; -.
CC EMBL: AF158101; AAD42595.1; -.
CC DR Hypothetical protein.
CC KW
CC SEQUENCE 64 AA; 7605 MW; 89E2AF66E86CCE0 CRC64;

Query Match 4.2%; Score 6; DB 1; Length 64;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 EKIKL 65
DB 8 EKIKL 13

RESULT 12
FER_SACER
ID FER_SACER STANDARD; PRT; 105 AA.
AC P24496;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Ferredoxin.
GN FDXA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=RM22;
RX MEDLINE=91276248; PubMed=2055472;
RA Donadio S., Hutchinson C.R.;
RT "Cloning and characterization of the Saccharopolyspora erythraea fdxa
RT gene encoding ferredoxin."
RN Gene 100:231-235(1991).
RN [2]
RP SEQUENCE OF 1-15.
RX MEDLINE=88169474; PubMed=3127376;
RA Shafiee A., Hutchinson C.R.;
RT "Purification and reconstitution of the electron transport components
RT for 6-deoxyerythronolide B hydroxylase, a cytochrome P-450 enzyme of
RT macrolide antibiotic (erythromycin) biosynthesis."
RN J. Bacteriol. 170:1548-1553(1988).
CC
CC -1- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -1- COFACTOR: BINDS 1 4FE-4S CLUSTER AND A 3PE-4S CLUSTER.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL TYPE FERREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M61119; AAA92023.1; -

DR PIR; JH0239; JH0239.

DR HSSP; Q45560; 1BD6.

DR InterPro; IPR001450; 4Fe4S-ferredoxin.

DR InterPro; IPR000813; 7Fe-ferredoxin.

DR Pfam; PF00037; Fer4; 1.

DR PRINTS; PR00353; 4FE4SFRDOXIN.

DR PRINTS; PR00354; 7FE8SFRDOXIN.

DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1.

KM Electron transport; Iron-sulfur; Duplication; 4Fe-4S; 3Fe-4S.

FT INIT\_MET 0 0

FT METAL 8 8 IRON-SULFUR 1 (3FE-4S) (BY SIMILARITY).

FT METAL 16 16 IRON-SULFUR 1 (3FE-4S) (BY SIMILARITY).

FT METAL 20 20 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).

FT METAL 39 39 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).

FT METAL 42 42 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).

FT METAL 45 45 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).

FT METAL 49 49 IRON-SULFUR 1 (3FE-4S) (BY SIMILARITY).

SO SEQUENCE 105 AA; 11407 MW; F42D85AC36406683 CRC64;

Query Match 4.2%; Score 6; DB 1; Length 105;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 VDRDVE 47  
|||||  
DB 90 VDRDVE 95

RESULT 13  
COX1\_SALTR STANDARD; PRT; 109 AA.

AC P29653;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).

GN COXI OR COI.

OS Salmo trutta (Brown trout).

OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;

OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.

OX NCBI\_TaxID=8032;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92130804; PubMed=1663569;

RA Normark B.B., McCune A.R., Harrison R.G.;

RT "Phylogenetic relationships of neopterygian fishes, inferred from mitochondrial DNA sequences.";

RL Mol. Biol. Evol. 8:819-834(1991).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B.

CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferri-cyclochrome c + 2 H(2)O.

CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M64917; AAB01484.1; -

DR HSSP; P00396; 1OCC.

DR InterPro; IPR000883; COX1.

DR Pfam; PF00115; COX1; 1.

DR PROSITE; PS00077; COX1; 1.

KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;

KW Respiratory chain; Inner membrane.

FT NON\_TER 1 1

FT METAL 6 6 COPPER B (PROBABLE).

FT METAL 10 10 COPPER B (PROBABLE).

FT METAL 56 56 COPPER B (PROBABLE).

FT METAL 57 57 COPPER B (PROBABLE).

FT NON\_TER 109 109

SO SEQUENCE 109 AA; 12251 MW; 0A513F2CE5B85C25 CRC64;

Query Match 4.2%; Score 6; DB 1; Length 109;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 IGLIGF 105  
|||||  
DB 46 IGLIGF 51

RESULT 14  
Y884\_CHLPN STANDARD; PRT; 117 AA.

AC Q92722; Q9JQ92;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein CPN0884/CP0982/CPJ0884.

GN CPN0884 OR CP0982 OR CPJ0884.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;

RX MEDLINE=99206606; PubMed=10192388;

RA Kalman S., Mitchell W., Marathe R., Jannet C., Fan J., Hyman R.W., Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Uppback T., Berry K., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwyn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=J138;

RX MEDLINE=20330349; PubMed=10871362;

RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;

RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";

RL Nucleic Acids Res. 28:2311-2314(2000).

CC -1- SIMILARITY: BELONGS TO THE UPF0092 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AEO01669; AAD19022.1; -;  
 DR EMBL: AEO02256; AAF38761.1; -;  
 DR EMBL: AP002548; BAA99092.1; -;  
 DR TIGR: CP0987; -;  
 DR InterPro: IPR003849; DUF219.  
 DR Pfam: PF02659; DUF219, 1.  
 DR Hypothetical protein; Transmembrane; Complete proteome.  
 KM TRANSMEM 30 50 POTENTIAL.  
 FT SEQUENCE 117 AA; 13152 MW; 45D3D3AC8B9E11A2 CRC64;

Query Match 4.28; Score 6; DB 1; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 RKNDLA 37  
 |||||  
 Db 63 RKNDLA 68

RESULT 15  
 LYC\_ANOGA STANDARD; PRT: 140 AA.  
 AC Q17005;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Lysozyme precursor (EC 3.2.1.17) (1/4-beta-N-acetylmuramidase).  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Anopheles.  
 OX NCBI\_TaxID=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97045819; PubMed=8890741;  
 RA Kang D., Romans P., Lee J.Y.;  
 RT "Analysis of a lysozyme gene from the malaria vector mosquito,  
 RT Anopheles gambiae.";  
 RL Gene 174:239-244(1996).  
 CC -I- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION; THOSE  
 CC IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-  
 CC MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOGENS.  
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-  
 CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan  
 CC heteropolymers of the prokaryotes cell walls.  
 CC -I- SIMILARITY: BELONGS TO FAMILY 22 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U28809; AAC47326.1; -;  
 DR HSSP: P00698; IAT5.  
 DR InterPro: IPR001916; Lactalbumin\_lysozyme.  
 DR Pfam: PF00062; lys; 1.  
 DR PRINTS: PRO0135; LYSLACT.  
 DR SMART: SM00263; LY21; 1.  
 DR PROSITE: PS00128; LACTALBUMIN\_LYSOZYME; 1.  
 KW Hydrolyase; Glycosidase; Bacteriolytic enzyme; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 140 LYSOZYME.

FT DISULFID 26 139 BY SIMILARITY.  
 FT DISULFID 47 128 BY SIMILARITY.  
 FT DISULFID 81 94 BY SIMILARITY.  
 FT DISULFID 90 108 BY SIMILARITY.  
 FT ACT\_SITE 52 52 BY SIMILARITY.  
 FT ACT\_SITE 69 69 BY SIMILARITY.  
 SQ SEQUENCE 140 AA; 15398 MW; 93AD614699216C28 CRC64;

Query Match 4.28; Score 6; DB 1; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 KNSSTD 23  
 |||||  
 Db 64 KNSSTD 69

Search completed: August 15, 2002, 11:24:40  
 Job time: 687 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:06 ; Search time 80.39 Seconds  
(without alignments)  
309,880 Million cell updates/sec

Title: US-08-981-087a-2  
Sequence: 144  
1 SYTNDKILILYFNKLYKKIK.....LNTNKIITWLTODTACGNOKL 144

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	100.0	1278	2	Q57236 Clostridium
2	26	18.1	1280	2	Q92A25 Clostridium
3	15	10.4	1268	2	Q45851 Clostridium
4	11	7.6	367	2	Q45862 Clostridium
5	11	7.6	367	2	Q45861 Clostridium
6	11	7.6	1251	2	Q9K395 Clostridium
7	11	7.6	1251	2	Q9FAR6 Clostridium
8	9	6.2	1296	2	Q45894 Clostridium
9	8	5.6	361	2	Q45846 Clostridium
10	8	5.6	361	2	Q45848 Clostridium
11	8	5.6	441	2	Q9X708 Clostridium
12	8	5.6	1072	16	Q9CF64 Clostridium
13	8	5.6	1291	2	Q92A28 Clostridium
14	8	5.6	1291	2	Q08077 Clostridium
15	8	5.6	1291	2	Q93G71 Clostridium
16	8	5.6	1291	2	Q933K0 Clostridium

17	7	4.9	53	13	Q90W19	Q90W19 gallus galli
18	7	4.9	111	13	Q90X55	Q90X55 gallus galli
19	7	4.9	122	13	Q90X56	Q90X56 gallus galli
20	7	4.9	169	9	Q9G0E2	Q9G0E2 lactococcus
21	7	4.9	177	16	Q91665	Q91665 borrelia bu
22	7	4.9	209	17	Q97VC3	Q97VC3 sulfolobus
23	7	4.9	241	17	Q97UQ7	Q97UQ7 sulfolobus
24	7	4.9	244	12	Q9YV06	Q9YV06 melanoplus
25	7	4.9	247	10	Q9KWM3	Q9KWM3 staphylococ
26	7	4.9	261	10	Q9SIF3	Q9SIF3 arabidopsis
27	7	4.9	330	16	Q99WV1	Q99WV1 staphylococ
28	7	4.9	448	2	Q9AKD7	Q9AKD7 rickettsia
29	7	4.9	458	5	Q9NAP3	Q9NAP3 caenorhabd
30	7	4.9	484	5	Q22967	Q22967 caenorhabd
31	7	4.9	648	3	Q9C469	Q9C469 schistosach
32	7	4.9	737	16	Q51274	Q51274 borrelia bu
33	7	4.9	842	2	Q93U59	Q93U59 candidatus
34	7	4.9	865	12	Q913V4	Q913V4 porcine gam
35	7	4.9	885	5	Q9BHY8	Q9BHY8 leishmania
36	7	4.9	890	10	Q93ZT8	Q93ZT8 arabidopsis
37	7	4.9	1010	5	Q18346	Q18346 drosophila
38	7	4.9	1037	5	Q9N9L3	Q9N9L3 leishmania
39	7	4.9	1116	5	Q19643	Q19643 caenorhabd
40	7	4.9	1711	5	Q96108	Q96108 plasmodium
41	7	4.9	2292	12	Q66763	Q66763 encephalomy
42	7	4.9	2292	12	Q89272	Q89272 encephalomy
43	6	4.2	52	2	Q985H1	Q985H1 clostridium
44	6	4.2	56	10	Q98975	Q98975 arabidopsis
45	6	4.2	57	10	Q98974	Q98974 arabidopsis

#### ALIGNMENTS

RESULT 1  
ID Q57236 PRELIMINARY; PRT: 1278 AA.  
AC Q57236; Q45863;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE BOTULINUM NEUROTOXIN TYPE F (BONT/F PROTEIN).  
GN BONT/F.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTCC 10281;  
RA Hutson R.A., Collins M.D.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 635-1000 FROM N.A.  
RC STRAIN=NTCC 1028;  
RX MEDLINE=94013372; PubMed=8408542;  
RA Campbell K., East A.K., Collins M.D.;  
RT "Gene probes for identification of the botulinum neurotoxin gene and specific identification of neurotoxin types B, E, and F".  
RL J. Clin. Microbiol. 31:2255-2262(1993).  
RN [4]  
RP SEQUENCE OF 1-37 FROM N.A.  
RC STRAIN=LANGE1AND;  
RX MEDLINE=98404102; PubMed=9732534;  
RA East A.K., Bhandari M., Hiem S., Collins M.D.;  
RT "Analysis of the botulinum neurotoxin type F gene clusters in proteolytic and nonproteolytic Clostridium botulinum and Clostridium baratii".  
RL Curr. Microbiol. 37:262-268(1998).

DR EMBL: X81714; CA57358.1; -;  
 DR EMBL: L35496; AAA23210.1; -;  
 DR EMBL: X70821; CAA50152.1; -;  
 DR EMBL: X99064; CAA67512.1; -;  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -;  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 DR Neurotoxin.  
 KM SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64;

Query Match 100.0%; Score 144; DB 2; Length 1278;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-135;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYTDKILILYFNKLYKKIKDNTSLDMRYENKFDISGYGNSISNGDVIYSTNRNQF 60  
 DB 848 SYTDKILILYFNKLYKKIKDNTSLDMRYENKFDISGYGNSISNGDVIYSTNRNQF 907  
 QY 61 GYSSKSEVNIQNDITNGRQNSISFWVRIPKFKENVNLNETTIIDCIRNNNSG 120  
 DB 908 GYSSKSEVNIQNDITNGRQNSISFWVRIPKFKENVNLNETTIIDCIRNNNSG 967  
 QY 121 WKISLNTKIIWTLQDTAGNOKL 144  
 DB 968 WKISLNTKIIWTLQDTAGNOKL 991

RESULT 2  
 Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.  
 ID Q9ZAJ5  
 AC Q9ZAJ5:  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE BONT PROTEIN.  
 GN BONT.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 3281 (ATCC 43757);  
 RC MEDLINE=98440323; PubMed=9767710;  
 RA Santos-Buelga J., Collins M.D., East A.K.;  
 RA "Characterization of the genes encoding the Botulinum neurotoxin  
 RT complex in a strain of clostridium botulinum producing type B & F  
 RT neurotoxins.";  
 RL Curr. Microbiol. 37:312-318(1998).  
 DR EMBL: Y18631; CAA73972.1; -;  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -;  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 DR SEQUENCE 1280 AA; 147487 MW; D0F748976BEC222C CRC64;

Query Match 18.1%; Score 26; DB 2; Length 1280;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-17;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 SILDREYENKFDISGYGNSISNG 48  
 ||||||||||||||||||||||||||||

DB 870 SILDREYENKFDISGYGNSISNG 895

RESULT 3  
 ID Q45851 PRELIMINARY; PRT; 1268 AA.  
 AC Q45851:  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE NEUROTOXIN TYPE F.  
 GN BONT /F.  
 OS Clostridium baratii.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1561;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=93252228; PubMed=8486245;  
 RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,  
 RA Richardson P.T.;  
 RT "Nucleotide sequence of the gene coding for Clostridium baratii type F  
 RT neurotoxin: Comparison with other clostridial neurotoxins.";  
 RL FEMS Microbiol. Lett. 108:175-182(1993).  
 DR EMBL: X68262; CAA48329.1; -;  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -;  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 DR SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 10.4%; Score 15; DB 2; Length 1268;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 51 YIYSTNRNQFGIYSS 65  
 DB 889 YIYSTNRNQFGIYSS 903

RESULT 4  
 ID Q45862 PRELIMINARY; PRT; 367 AA.  
 AC Q45862:  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
 DE BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).  
 GN BONT/E.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TYPE E. HAZEN 36208 (ATCC 9564);  
 RC MEDLINE=94013372; PubMed=8408542;  
 RA Campbell K., East A.K., Collins M.D.;  
 RA "Gene probes for identification of the botulinal neurotoxin gene and  
 RT specific identification of neurotoxin types B, E, and F.";  
 RL J. Clin. Microbiol. 31:2255-2262(1993).  
 DR EMBL: X70815; CAA50146.1; -;  
 DR HSSP: P10845; 3BTA.  
 DR Neurotoxin.  
 KW NON\_TER 1  
 FT NON\_TER 367  
 SQ SEQUENCE 367 AA; 42854 MW; 0810595B3A865570 CRC64;

	Query Match	7.6%;	Score 11;	DB 2;	Length 367;	
	Best Local Similarity	100.0%;	Pred. No. 0.0072;			
Matches	11; Conservative	0;	Mismatches	0;	Indels	0;
OY	86 NFSISFWVRIP 96 				Gaps	
Db	299 NFSISFWVRIP 309					
RESULT	5					
ID	Q45861	PRELIMINARY;	PRT:	367 AA.		
AC	Q45861;					
DT	01-NOV-1996 (TREMBLER). 01, Created)					
DR	01-NOV-1996 (TREMBLER). 01, last sequence update)					
DE	01-OCT-2000 (TREMBLER). 15, last annotation update)					
GN	BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).					
OS	Clostridium botulinum.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;					
CC	Clostridium.					
NCBI_TaxID	-1491;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-TYPE E, VH (DOLAN);					
RA	MEDLINE=94013372; PubMed=8408542;					
RT	"Campbell K., East A.K., Collins M.D.;					
RL	gene probes for identification of the botulin neurotoxin gene and					
DR	J. Clin. Microbiol. 31:2255-2262(1993).					
EMBL	X70818; CAA50149.1; -					
HSSP	P10845; 3BTA					
KW	Neurotoxin.					
FT	NON_TER 1 1					
FT	NON_TER 367 367					
SQ	SEQUENCE 367 AA; 42902 MW; 346A610C2FF70262 CRC64;					
Query Match	7.6%;	Score 11;	DB 2;	Length 367;		
Best Local Similarity	100.0%;	Pred. No. 0.0072;				
Matches	11; Conservative	0;	Mismatches	0;	Indels	0;
OY	86 NFSISFWVRIP 96 				Gaps	
Db	299 NFSISFWVRIP 309					
RESULT	6					
OYK395	PRELIMINARY;	PRT:	1251 AA.			
ID	Q9K395;					
AC	Q9K395;					
DT	01-OCT-2000 (TREMBLER). 15, Created)					
DR	01-OCT-2000 (TREMBLER). 15, last sequence update)					
DE	01-DEC-2001 (TREMBLER). 19, last annotation update)					
DE	TYPE E BOTULINUM TOXIN.					
GN	BONT/E.					
OS	Clostridium butyricum.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;					
CC	Clostridium.					
NCBI_TaxID	-1492;					
OX	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-LCL 095;					
RA	Wang X., Megawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,					
RA	Karasawa T.,					
RT	C. butyricum (LCL 095) gene for type E botulinum toxin.";					
RU	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-LCL 155 (KZ 1885);					
RA	Wang X., Megawa T., Kozaki S., Tsukamoto K., Gyodo Y., Yamakawa K.,					
RA	Kato H., Nakamura S., Karasawa T.,					

[illegible]

DR EMBL: AB037711; BAB03519.1; -  
 DR EMBL: AB037712; BAB03520.1; -  
 DR EMBL: AB037713; BAB03521.1; -  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN; 1.  
 SQ SEQUENCE 1251 AA; 143751 MW; 2021F4E427070296 CRC64;

Query Match 7.6%; Score 11; DB 2; Length 1251;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 NFSISFWVRIP 96  
 |||||

DB 914 NFSISFWVRIP 924

RESULT 7  
 O9FAR6 PRELIMINARY; PRT; 1255 AA.  
 AC O9FAR6;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TYPE E BOTULINUM TOXIN.  
 GN BONT/E.  
 OS Clostridium butyricum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1492;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BL 6340/ATCC 43755/BL 5520/KZ 147;  
 RX MEDLINE=20509829; PubMed=11055954;  
 RA Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K., Oguma K., Sakauchi Y., Nakamura S.;  
 RT "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium butyricum Strains";  
 RL Appl. Environ. Microbiol. 66:4992-4997(2000).  
 DR EMBL: AB039264; BAB12249.1; -  
 DR HSSP: P10845; 3BTA.  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN; 1.  
 SQ SEQUENCE 1255 AA; 143918 MW; 1B557B9D85CDBE4D CRC64;

Query Match 7.6%; Score 11; DB 2; Length 1255;  
 Best Local Similarity 100.0%; Pred. No. 0.018;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 NFSISFWVRIP 96  
 |||||

DB 917 NFSISFWVRIP 927

RESULT 8  
 Q45894 PRELIMINARY; PRT; 1296 AA.  
 ID Q45894;  
 AC Q45894; P77780;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE BOTULINUM NEUROTOXIN TYPE A (TYPE A NEUROTOXIN).

GN BONT OR ATX.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KYOTO-F;  
 RX MEDLINE=94143603; PubMed=8310180;  
 RA Williams A., East A.K., Lawson P.A., Collins M.D.;  
 RT "Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins";  
 RL Res. Microbiol. 144:547-556(1993).  
 RN [2]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RC STRAIN-62A;  
 RX MEDLINE=97016817; PubMed=8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RT "Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic Clostridium botulinum types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 RN [3]  
 RP SEQUENCE OF 1-18 FROM N.A.  
 RC STRAIN-TYPE A NIH;  
 RX MEDLINE=96096783; PubMed=8521962;  
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
 RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin components of Clostridium botulinum type A progenitor toxins";  
 RL FEBS Lett. 376:41-44(1995).  
 DR EMBL: X73423; CAA51824.1; -  
 DR EMBL: X92973; CAA63551.1; -  
 DR EMBL: X87974; CAA61234.1; -  
 DR EMBL: D67030; BAA11051.1; -  
 DR HSSP: P10845; 3BTA.  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 KW Neurotoxin.  
 SQ SEQUENCE 1296 AA; 149410 MW; 6F12E7BF28DE69D1 CRC64;

Query Match 6.2%; Score 9; DB 2; Length 1296;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 NLNNEYTII 111  
 |||||

DB 957 NLNNEYTII 965

RESULT 9  
 Q45846 PRELIMINARY; PRT; 361 AA.  
 ID Q45846;  
 AC Q45846;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).  
 GN BONT/B.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TYPE B, NON-PROTEOLYTIC 2129B (SCOTT);  
 RX MEDLINE=94013372; PubMed=8408542;  
 RA Campbell K., East A.K., Collins M.D.;  
 RT "Gene probes for identification of the botulinum neurotoxin gene and



```

RT Specific identification of neurotoxin types B, E, and F."
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL: X70814; CAA50145.1; -.
DR HSSP: P10845; 3BTA.
KM Neurotoxin.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42175 MW; 533EA98735CD98E1 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 NNSGWKIS 124
DB 325 NNSGWKIS 332

RESULT 10
ID 045848 PRELIMINARY; PRT; 361 AA.
AC 045848.
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE B, NON-PROTEOLYTIC EKLUND 2B (COLMORTH 229);
RA MEDLINE-9401372; PubMed-8408542;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F."
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL: X70815; CAA50150.1; -.
DR HSSP: P10845; 3BTA.
KM Neurotoxin.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42131 MW; A2E0FFFC81F9533D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 NNSGWKIS 124
DB 325 NNSGWKIS 332

RESULT 11
ID 09X708 PRELIMINARY; PRT; 441 AA.
AC 09X708.
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99343691; PubMed-10413679;

```

```

RA Talli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA Schiavo G.,
RT "Functional characterisation of tetanus and botulinum neurotoxins
RT binding domains".
RL J. Cell Sci. 112:2715-2724(1999).
DR EMBL: AJ242628; CAB43706.1; -.
DR HSSP: P10845; 3BTA.
KM Neurotoxin.
FT NON_TER 1 1
FT NON_TER 441 441
SQ SEQUENCE 441 AA; 52772 MW; 721DOB468EC95A4 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 NNSGWKIS 124
DB 116 NNSGWKIS 123

RESULT 12
ID 09CF64 PRELIMINARY; PRT; 1072 AA.
AC 09CF64.
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
DE UNKNOWN PROTEIN.
GN YOF.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IL1403;
RA MEDLINE-21235186; PubMed-11337471;
RX Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Wiesenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006392; AAK05715.1; -.
KM Complete Proteome.
SQ SEQUENCE 1072 AA; 113056 MW; 464446E2656CA08 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 SSKPSEVN 71
DB 712 SSKPSEVN 719

RESULT 13
ID 09ZAJ8 PRELIMINARY; PRT; 1291 AA.
AC 09ZAJ8.
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE BONT. PROTEIN.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-CDC 3281 (ATCC 43757);  
 RA MEDLINE=98440323; PubMed=9767710;  
 RA Santos-Buelga J., Collins M.D., East A.K.;  
 RT "Characterization of the genes encoding the Botulinum neurotoxin  
 RT complex in a strain of Clostridium botulinum producing type B & F  
 RT neurotoxins";  
 RL Curr. Microbiol. 37:312-318(1998).  
 DR EMBL: Y13630; CAA73968.1; -.  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -.  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27.1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 DR SEQUENCE 1291 AA; 150840 MW; EAD3B0E46AB2E735 CRC64;

Query Match 5.6%; Score 8; DB 2; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 NNSGKITS 124  
 |||||||  
 DB 958 NNSGKITS 965

## RESULT 14

ID 008077 PRELIMINARY; PRT; 1291 AA.

AC 008077;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE BOTULINUM NEUROTOXIN TYPE B (EC 3.4.24.-) (BONT/B).  
 GN BONT/B.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EKUND 17B ATCC25765;  
 RX MEDLINE=94122659; PubMed=7764370;  
 RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;  
 RT "Nucleotide sequence of the gene coding for non-proteolytic  
 RT Clostridium botulinum type B neurotoxin: comparison with other  
 RT Clostridial neurotoxins";  
 RL Curr. Microbiol. 28:101-110(1994).  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CLEAVES SYNAPTOSOMAL-2.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: HIGH WITH OTHER BOTULINUM NEUROTOXINS AND WITH TETANUS  
 CC NEUROTOXIN.  
 CC -1- SIMILARITY: TO OTHER ZINC METALLOPROTEINASES IN THE ACTIVE SITE  
 CC REGION  
 DR EMBL: X71343; CAA50482.1; -.  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27.002; -.  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01742; Peptidase\_M27.1.

DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAA CRC64;

Query Match 5.6%; Score 8; DB 2; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 NNSGKITS 124  
 |||||||  
 DB 958 NNSGKITS 965

RESULT 15  
 ID 093671 PRELIMINARY; PRT; 1291 AA.

AC 093671;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE NEUROTOXIN TYPE B.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1436;  
 RA Kirma N., Ferreira J.L., Baumstark B.R.;  
 RT "Characterization of six type A strains of Clostridium botulinum that  
 RT contain type B toxin gene sequences";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF295926; AAK97132.1; -.  
 DR SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 5.6%; Score 8; DB 2; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 NNSGKITS 124  
 |||||||  
 DB 958 NNSGKITS 965

Search completed: August 15, 2002, 11:24:07  
 Job time: 694 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:07 ; Search time 80.39 Seconds  
(without alignments) 309.880 Million cell updates/sec

Title: US-08-981-087a-3

Sequence: 1 VENTOMISIDYINKMIFV.....ITQNSMFLINQRCGYOKP 144

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL\_19.\*

1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_viral:\*  
16: sp.\_bacteriophage:\*  
17: sp.\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	144	100.0	1278	2	057236 clostridium
2	31	21.5	1280	2	092A55
3	22	15.3	1268	2	045851
4	15	10.4	1251	2	09K395
5	15	10.4	1255	2	09FAR6
6	11	7.6	1296	2	045894
7	8	5.6	1296	2	045894
8	7	4.9	131	4	09B028
9	7	4.9	154	11	062080
10	7	4.9	166	4	095023
11	7	4.9	182	15	085641
12	7	4.9	187	15	085640
13	7	4.9	187	15	083401
14	7	4.9	200	16	09PMH8
15	7	4.9	202	11	062082
16	7	4.9	203	17	0979N7

17	7	4.9	231	2	046032	046032 clostridium
18	7	4.9	232	2	046027	046027 clostridium
19	7	4.9	263	2	093J58	093J58 neisseria m
20	7	4.9	263	2	093J56	093J56 neisseria l
21	7	4.9	267	16	093X73	093X73 neisseria m
22	7	4.9	307	16	09HXW8	09HXW8 pseudomonas
23	7	4.9	358	12	09MR17	09MR17 african cas
24	7	4.9	358	12	09JEA2	09JEA2 cassava gem
25	7	4.9	358	12	09JEA2	09JEA2 cassava gem
26	7	4.9	379	17	029504	029504 archaeoglob
27	7	4.9	436	10	09FV25	09FV25 cryza saliv
28	7	4.9	441	2	09X708	09X708 clostridium
29	7	4.9	450	12	09BMO0	09BMO0 amesara mco
30	7	4.9	451	16	09TE21	09TE21 clostridium
31	7	4.9	534	5	09G711	09G711 caenorhabdit
32	7	4.9	636	15	085506	085506 murine leuk
33	7	4.9	669	15	09Y753	09Y753 murine leuk
34	7	4.9	814	10	022695	022695 cryza saliv
35	7	4.9	846	10	09FEP9	09FEP9 cryza saliv
36	7	4.9	1023	5	062398	062398 caenorhabdit
37	7	4.9	1036	11	091YD5	091YD5 mus musculu
38	7	4.9	1270	5	019736	019736 caenorhabdit
39	7	4.9	1272	12	010243	010243 clover yell
40	7	4.9	1275	12	09GTG7	09GTG7 clostridium
41	7	4.9	1280	2	09LBS7	09LBS7 clostridium
42	7	4.9	1280	2	045849	045849 clostridium
43	7	4.9	1291	2	092A58	092A58 clostridium
44	7	4.9	1291	2	008077	008077 clostridium
45	7	4.9	1291	2	093671	093671 clostridium

## ALIGNMENTS

RESULT 1  
ID 057236 PRELIMINARY; PRT; 1278 AA.  
AC 057236: 045863;  
DT 01-NOV-1996 (TREMBL) 01, Created  
DT 01-NOV-1996 (TREMBL) 01, Last sequence update  
DT 01-JUN-2001 (TREMBL) 17, Last annotation update  
DE BOTULINUM NEUROTOXIN TYPE F (BONT/F PROTEIN).  
GN BONT/F.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=NCTC 10281;  
RC Hutson R.A., Collins M.D.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Elmore M.J., Bodsworth N.J., Whelan G.M., Minton N.P.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 635-1000 FROM N.A.  
RA STRAIN=NCTC 1028;  
RC MEDLINE=94013372; PubMed=8408542;  
RA Campbell K., East A.K., Collins M.D.;  
RL "Gene probes for identification of the botulin neurotoxin gene and specific identification of neurotoxin types B, E, and F.";  
RN [4]  
RP Cln. Microbiol. 31:2255-2262(1993).  
RN [5]  
RP SEQUENCE OF 1-27 FROM N.A.  
RA STRAIN=LANGELEND.  
RC MEDLINE=9304102; PubMed=9732534;  
RA East A.K., Bhandari M., Helm S., Collins M.D.;  
RL "Analysis of the botulin neurotoxin type F gene clusters in proteolytic and nonproteolytic Clostridium botulinum and Clostridium baratti".  
RN Curr. Microbiol. 37:262-268(1998).





OC Clostridium.  
 RN NCBI\_TaxID=1491;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KYOTO-F;  
 RX MEDLINE=94143603; PubMed=8310180;  
 RA Williams A., East A.K., Lawson P.A., Collins M.D.;  
 RT "Sequence of the gene coding for the neurotoxin of Clostridium  
 RT botulinum type A associated with infant botulism: comparison with  
 RT other clostridial neurotoxins.";  
 RL Res. Microbiol. 144:547-556(1993).  
 RN [2]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RC STRAIN=62A;  
 RX MEDLINE=97016817; PubMed=8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RT "Organization and phylogenetic interrelationships of genes encoding  
 RT components of the botulinum toxin complex in proteolytic Clostridium  
 RT botulinum types A, B, and F: evidence of chimeric sequences in the  
 RT gene encoding the nontoxic nonhemagglutinin component.";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 RN [3]  
 RP SEQUENCE OF 1-18 FROM N.A.  
 RC STRAIN=TYPE A NIH;  
 RX MEDLINE=96096783; PubMed=8521962;  
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
 RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
 RT components of Clostridium botulinum type A progenitor toxins.";  
 RL FBS Lett. 376:41-44(1995).  
 DR EMBL: X73423; CAA51824.1; -;  
 DR EMBL: X92973; CAA63551.1; -;  
 DR EMBL: X87974; CAA61234.1; -;  
 DR EMBL: D67030; BA11051.1; -;  
 DR HSSP: P10845; 3BTA.  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOXILYSIN.  
 DR ProDom: PD001963; Bontoxilysin; 1.  
 KW Neurotoxin.  
 SQ SEQUENCE 1296 AA; 149410 MW; 6F12E7BF28ED69D1 CRC64;

Query Match 7.6%; Score 11; DB 2; Length 1296;  
 Best Local Similarity 100.0%; Pred. No. 0.015;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 17 WIFVTINRL 27  
 |||||||||  
 DB 1014 WIFVTINRL 1024

RESULT 7  
 ID 09A928 PRELIMINARY; PRT; 540 AA.  
 AC 09A928;  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE HYPOTHEICAL PROTEIN CC0813.  
 GN CC0813.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11559647;  
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Debey R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Knouri H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL: AE005758; AAK22798.1; -;  
 DR TIGR: CC0813; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 540 AA; 59648 MW; 72BC45442BEP99FD CRC64;

Query Match 5.6%; Score 8; DB 16; Length 540;  
 Best Local Similarity 100.0%; Pred. No. 8.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 94 PDPSTLKD 101  
 |||||||  
 DB 71 PDPSTLKD 78

RESULT 8  
 ID 09BUA6 PRELIMINARY; PRT; 131 AA.  
 AC 09BUA6;  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE SIMILAR TO MYOSIN LIGHT CHAIN 2, PRECURSOR LYMPHOCTE-SPECIFIC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MELANOMA.;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC002778; AA02778.1; -;  
 DR HSSP: P13543; ISCM.  
 DR InterPro: IPR002048; EF-hand.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN; 1.  
 SQ SEQUENCE 131 AA; 14930 MW; 336C55E3C70C07A CRC64;

Query Match 4.9%; Score 7; DB 4; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 74 FKVPDTE 80  
 |||||||  
 DB 68 FKVPDTE 74

RESULT 9  
 ID 062080 PRELIMINARY; PRT; 154 AA.  
 AC 062080;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE MYOSIN LIGHT CHAIN 2 (FRAGMENT).  
 GN MYLC2PL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92331628; PubMed=1628631;  
 RA Oltz E.M., Yancopoulos G.D., Morrow M.A., Rolink A., Lee G., Wong F.,  
 RA Kaplan K., Gillis S., Melchers F., Alt F.W.;  
 RT "A novel regulatory myosin light chain gene distinguishes pre-B cell  
 RT subsets and is IL-7 inducible.";  
 RL EMBO J. 11:2759-2767(1992).



DR EMBL: X65979; CAA46794.1; -.  
 DR HSSP: P13543; 1SCM.  
 DR MGI: MGI:1891703; Mylczpl.  
 FT NON\_TER 1  
 SQ SEQUENCE 154 AA; 17439 MW; 085265DBAE42912F CRC64;

Query Match 4.9%; Score 7; DB 11; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 FKVPDTE 80  
 |||||  
 DB 91 FKVPDTE 97

RESULT 10  
 095023 PRELIMINARY; PRT; 166 AA.

ID 095023;  
 AC 095023;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DE MUGSC-H.D11059M17.2 PROTEIN (FRAGMENT).  
 GN MUGSC-H.D11059M17.2  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Geisel C., Kallck J., Gibson A.;  
 RT The sequence of Homo sapiens PAC clone RP5-1059M17.2;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
 DR EMBL: AC004953; AAD08850.1; -.  
 DR HSSP: P13543; 1SCM.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; efhand; 2.  
 DR SMART: SM00054; EFh; 2.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
 KW Calcium-binding.  
 FT NON\_TER 1  
 SQ SEQUENCE 166 AA; 18917 MW; BCD43CB94A931605 CRC64;

Query Match 4.9%; Score 7; DB 4; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 FKVPDTE 80  
 |||||  
 DB 103 FKVPDTE 109

RESULT 11  
 085641

ID 085641; PRELIMINARY; PRT; 182 AA.

AC 085641;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE 3' END OF THE GENOME OF MOLONEY MURINE LEUKEMIA VIRUS (CODES FOR THE ENV GENE) (FRAGMENT).  
 OS Moloney murine leukemia virus.  
 OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus.  
 OX NCBI\_TaxID=11801;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE-81013872; PubMed-6251454;  
 RA Sutcliffe J.G., Shinnick T.M., Verma I.M., Lerner R.A.;  
 RT "Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of replicatons, analogy to bacterial transposons, and an unexpected gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:3302-3306(1980).  
 DR EMBL: V01178; CAA24501.1; -.  
 DR HSSP: P03385; 1MOF.  
 DR InterPro: IPR002050; Env\_polyprotein.  
 DR Pfam: PF00429; ENV\_polyprotein; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 182 AA; 20384 MW; 9212B2B384E97328 CRC64;

Query Match 4.9%; Score 7; DB 15; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 EKISNLT 47  
 |||||  
 DB 37 EKISNLT 43

RESULT 12  
 085640 PRELIMINARY; PRT; 187 AA.

ID 085640;  
 AC 085640;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE MOLONEY MURINE LEUKEMIA VIRUS (M-MULV) TRANSPON GENES AND DEPLICATING DETAILS (FRAGMENT).  
 OS Murine leukemia virus.  
 OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus.  
 OX NCBI\_TaxID=11786;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-81259595; PubMed-626762;  
 RA Sutcliffe J.G., Shinnick T.M., Lerner R.A.;  
 RT "Moloney murine leukemia virus is a transposon: Nucleotide sequence analysis identifies genes and replication details.";  
 RL Cold Spring Harb. Symp. Quant. Biol. 45:707-710(1981).  
 DR EMBL: M12997; AAA46529.1; -.  
 DR HSSP: P03385; 1MOF.  
 DR InterPro: IPR002050; Env\_polyprotein.  
 DR Pfam: PF00429; ENV\_polyprotein; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 187 AA; 20841 MW; C0832DDEFA8D7CA5 CRC64;

Query Match 4.9%; Score 7; DB 15; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 EKISNLT 47  
 |||||  
 DB 42 EKISNLT 48

RESULT 13  
 083401

ID 083401; PRELIMINARY; PRT; 187 AA.

AC 083401;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE POLYPROTEIN PRECURSOR (FRAGMENT).  
 OS Moloney murine sarcoma virus.  
 OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus.  
 OX NCBI\_TaxID=11809;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-81052384; PubMed-6159543;

RA Sutcliffe J.G., Shinnick T.M., Green N., Liu F.T., Niman H.L.,  
 RA Lerner R.A.;  
 RT "Chemical synthesis of a polypeptide predicted from nucleotide  
 RT sequence allows detection of a new retroviral gene product.";  
 RL Nature 287:801-805(1980).  
 RN [2]  
 RP SEQUENCE OF 6-187 FROM N.A.  
 RX MEDLINE=81013872; PubMed=6251454;  
 RA Sutcliffe J.G., Shinnick T.M., Verna I.M., Lerner R.A.;  
 RT "Nucleotide sequence of Moloney leukemia virus: 3' end reveals details  
 RT of replications, analogy to bacterial transposons, and an unexpected  
 RT gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:3302-3306(1980).  
 DR EMBL: J02261; AA51623.1; -.  
 DR HSSP: P03385; IMOP.  
 DR InterPro: IPR002050; Env\_polyprotein.  
 DR Pfam: PF00429; Env\_polyprotein; 1.  
 FT NON\_TER 1 1  
 FT CHAIN 1 94 POTENTIAL.  
 FT 95 187 POTENTIAL.  
 SQ SEQUENCE 187 AA; 20841 MW; C0832DDE4BD7CA5 CRC64;

Query Match 4.9%; Score 7; DB 15; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 EKSISNL 47  
 |||||  
 Db 42 EKSISNL 48

RESULT 14  
 O9PMH8 PRELIMINARY; PRT; 200 AA.  
 AC O9PMH8;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE PUTATIVE MEMBRANE PROTEIN.  
 GN CJI484C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxId=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrell B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 DR EMBL: AL139078; CAB73906.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 200 AA; 22387 MW; 753C07F428DA45BE CRC64;

Query Match 4.9%; Score 7; DB 16; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 KTEIETL 89  
 |||||  
 Db 66 KTEIETL 72

RESULT 15  
 Q62082

ID 062082 PRELIMINARY; PRT; 202 AA.  
 AC 062082;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE MYOSIN LIGHT CHAIN 2.  
 GN MYLC2PL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92331628; PubMed=1628631;  
 RA Oltz E.M., Yancopoulos G.D., Morrow M.A., Rolink A., Lee G., Wong F.,  
 RA Kaplan K., Gillis S., Melchers F., Alt F.W.;  
 RT "A novel regulatory myosin light chain gene distinguishes pre-B cell  
 RT subsets and is IL-7 inducible.";  
 RL EMBO J. 11:2759-2767(1992).  
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.  
 DR EMBL: X65981; CAA46796.1; -.  
 DR HSSP: P13543; ISCM.  
 DR MGD; MGI:1891705; Mylc2pl.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; efhand; 2.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
 KW Calcium-binding.  
 SQ SEQUENCE 202 AA; 22511 MW; 97C976AD9E1A90BD CRC64;

Query Match 4.9%; Score 7; DB 11; Length 202;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 FKVPDTE 80  
 |||||  
 Db 139 FKVPDTE 145

Search completed: August 15, 2002, 11:24:08  
 Job time: 695 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:08 ; Search time 80.39 Seconds  
(without alignments)  
307.728 Million cell updates/sec

Title: US-08-981-087a-4

Perfect score: 143  
Sequence: 1 NIESNRLTYGVEVIRKNG.....TSSNGCFWSFKSEHMOEN 143

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size: 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

- 1: SP\_RMBL\_19:\*\*
- 2: SP\_Archaea:\*\*
- 3: SP\_Bacteria:\*\*
- 4: SP\_Fungi:\*\*
- 5: SP\_Human:\*\*
- 6: SP\_Invertebrate:\*\*
- 7: SP\_Mammal:\*\*
- 8: SP\_Mhc:\*\*
- 9: SP\_Organelle:\*\*
- 10: SP\_Phage:\*\*
- 11: SP\_Plant:\*\*
- 12: SP\_Rodent:\*\*
- 13: SP\_Virus:\*\*
- 14: SP\_Unclassified:\*\*
- 15: SP\_Yrirus:\*\*
- 16: SP\_Bacteriap:\*\*
- 17: SP\_Archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	100.0	1278	2	057236 Clostridium
2	22	15.4	1280	2	092A35 Clostridium
3	16	11.2	1268	2	045851 Clostridium
4	8	5.6	96	15	0994N3 human immun
5	8	5.6	96	15	0994N3 human immun
6	8	5.6	731	5	0994N3 human immun
7	8	5.6	1251	2	0994N3 human immun
8	8	5.6	1251	2	0994N3 human immun
9	8	4.9	67	16	092C93 Clostridium
10	7	4.9	96	15	042077 human immun
11	7	4.9	96	15	089653 human immun
12	7	4.9	96	15	089654 human immun
13	7	4.9	96	15	089670 human immun
14	7	4.9	96	15	089677 human immun
15	7	4.9	96	15	089678 human immun
16	7	4.9	96	15	089679 human immun

17	7	4.9	96	15	090008	090qg8 human immun
18	7	4.9	96	15	090029	090d29 human immun
19	7	4.9	180	15	0908P6	090d29 human immun
20	7	4.9	242	3	090829	090829 candida alb
21	7	4.9	242	5	062519	062519 caenorhabd
22	7	4.9	249	8	095C19	095c19 sarcopera r
23	7	4.9	252	8	095C17	095c17 schwartzia
24	7	4.9	252	8	095C16	095c16 souroubea e
25	7	4.9	255	8	095C12	095c12 margravias
26	7	4.9	255	8	095C10	095c10 ruyachna ph
27	7	4.9	255	8	095C18	095c18 sarcopera s
28	7	4.9	255	8	095C15	095c15 sarcopera s
29	7	4.9	255	8	095C14	095c14 souroubea s
30	7	4.9	259	2	0969A4	0969a4 listeria in
31	7	4.9	259	2	085017	085017 listeria in
32	7	4.9	259	16	092DC1	092dc1 listeria in
33	7	4.9	264	16	051709	051709 borrelia bu
34	7	4.9	264	16	085743	085743 listeria mo
35	7	4.9	280	10	09M391	09m391 arabidopsis
36	7	4.9	347	5	062107	062107 caenorhabd
37	7	4.9	375	16	091769	091769 pseudomonas
38	7	4.9	380	16	09WYS8	09wys8 thermotoga
39	7	4.9	390	5	0950U5	095qu5 caenorhabd
40	7	4.9	396	16	034629	034629 bacillus su
41	7	4.9	407	17	09UXD3	09uxd3 sulfolobus
42	7	4.9	516	5	024282	024282 drosophila
43	7	4.9	518	5	09YU04	09yug4 drosophila
44	7	4.9	543	3	094342	094342 schizosacch
45	7	4.9	753	5	09X2J5	09x2j5 strongyloid

## ALIGNMENTS

RESULT	1	PREDIMINARY:	PRT:	1278 AA.
ID	057236	PRELIMINARY:		
AC	057236; 045863;			
DT	01-NOV-1996 (TREMURel. 01, Created)			
DT	01-NOV-1996 (TREMURel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMURel. 17, Last annotation update)			
DE	BOTULINUM NEUROTOXIN TYPE F (BONT/F PROTEIN).			
GN	BONT/F.			
OS	Clostridium botulinum.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OX	NCBI_TaxId=1491;			
RI	[1]			
RE	SEQUENCE FROM N.A.			
RC	STRAIN=NCIC 10281;			
RA	Hutson R.A., Collins M.D.;			
RL	Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;			
RL	Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 635-1000 FROM N.A.			
RC	STRAIN=NCIC 10281;			
RA	MEDLINE=94013372; PubMed=8408542;			
RT	Campbell K., East A.K., Collins M.D.;			
RT	"Gene probes for identification of the botulinum neurotoxin gene and			
RT	specific identification of neurotoxin types B, E, and F.";			
RT	J. Clin. Microbiol. 31:2255-2262(1993).			
RP	SEQUENCE OF 1-27 FROM N.A.			
RC	STRAIN=LANCELANO;			
RA	MEDLINE=98061102; PubMed=9742534;			
RT	East A.K., Bhandari M., Hielm S., Collins M.D.;			
RT	"Analysis of the botulinum neurotoxin type F gene clusters in			
RT	proteolytic and nonproteolytic Clostridium botulinum and Clostridium			
RT	bartell.";			
RT	Curr. Microbiol. 37:262-268(1998).			

DR EMBL: X81714; CAA57358.1; -;  
DR EMBL: L35496; AAA23210.1; -;  
DR EMBL: X70821; CAA50152.1; -;  
DR EMBL: X99064; CAA67512.1; -;  
DR HSSP: P10845; 3BTA.  
DR MEROPS: M27.002; -;  
DR InterPro: IPR000395; Bontoxilysin.  
DR InterPro: IPR000130; Zn\_MTPeptidse.  
DR Pfam: PF01742; Peptidase\_M27; 1.  
DR PRINTS: PR00760; BONTOXILYSIN.  
DR ProDom: PD001963; Bontoxilysin; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KM Neurotoxin.  
SQ SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64;

Query Match 100.0%; Score 143; DB 2; Length 1278;  
Best Local Similarity 100.0%; Pred. No. 2e-142;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NIFSNTRLYTGVEYIRKNGSTDISNTDNFVRKNDLAYINVVDREYRLYADISIAKPE 60  
DB 1136 NIFSNTRLYTGVEYIRKNGSTDISNTDNFVRKNDLAYINVVDREYRLYADISIAKPE 1195  
QY 61 KIILIRTSNNSNSLGGIIVVDSIGNNCTMNFQNNNGSINIGLGFHSNNLYASSWYNNI 120  
DB 1196 KIILIRTSNNSNSLGGIIVVDSIGNNCTMNFQNNNGSINIGLGFHSNNLYASSWYNNI 1255  
QY 121 RKNTSSNCGFWSFKSEHGQEN 143  
DB 1256 RKNTSSNCGFWSFKSEHGQEN 1278

RESULT 2  
Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.  
ID Q9ZAJ5  
AC Q9ZAJ5;  
DT 01-MAY-1999 (TREMblrel. 10, Created)  
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE BONT PROTEIN.  
GN BONT.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 3281 (ATCC 43757);  
RX MEDLINE=98440323; PubMed=9767710;  
RA Santos-Buelga J., Collins M.D., East A.K.;  
RT "Characterization of the genes encoding the Botulinum neurotoxin  
RT complex in a strain of clostridium botulinum producing type B & F  
RT neurotoxins.";  
RL Curr. Microbiol. 37:312-318(1998).  
DR EMBL: Y13631; CAA73972.1; -;  
DR HSSP: P10845; 3BTA.  
DR MEROPS: M27.002; -;  
DR InterPro: IPR000395; Bontoxilysin.  
DR InterPro: IPR000130; Zn\_MTPeptidse.  
DR Pfam: PF01742; Peptidase\_M27; 1.  
DR PRINTS: PR00760; BONTOXILYSIN.  
DR ProDom: PD001963; Bontoxilysin; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 1280 AA; 147487 MW; D0F748976B8C222C CRC64;

Query Match 15.4%; Score 22; DB 2; Length 1280;  
Best Local Similarity 100.0%; Pred. No. 1.9e-14;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 SLGGIIVVDSIGNNCTMNFQNN 95  
|||||

DB 1212 SLGGIIVVDSIGNNCTMNFQNN 1233

RESULT 3  
ID Q45851 PRELIMINARY; PRT; 1268 AA.  
AC Q45851;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE BONTROXIN TYPE F.  
GN BONT /F.  
OS Clostridium baratii.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1561;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9325228; PubMed=8486245;  
RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,  
RA Richardson P.T.;  
RT "Nucleotide sequence of the gene coding for Clostridium baratii type F  
RT neurotoxin: Comparison with other clostridial neurotoxins.";  
RL FEBS Microbiol. Lett. 108:175-182(1993).  
DR EMBL: X68262; CAA48529.1; -;  
DR HSSP: P10845; 3BTA.  
DR MEROPS: M27.002; -;  
DR InterPro: IPR000395; Bontoxilysin.  
DR InterPro: IPR000130; Zn\_MTPeptidse.  
DR Pfam: PF01742; Peptidase\_M27; 1.  
DR PRINTS: PR00760; BONTOXILYSIN.  
DR ProDom: PD001963; Bontoxilysin; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 11.2%; Score 16; DB 2; Length 1268;  
Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 127 NCGFWSFKSEHGQOE 142  
DB 1253 NCGFWSFKSEHGQOE 1268

RESULT 4  
ID Q994N3 PRELIMINARY; PRT; 96 AA.  
AC Q994N3;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)  
DE VPR PROTEIN.  
GN VPR.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=97ZA012;  
RX MEDLINE=21094715; PubMed=11177395;  
RA Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,  
RA Kalish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;  
RT "Near full-length clones and reference sequences for subtype C  
RT isolates for HIV type 1 from three different continents.";  
RL AIDS Res. Hum. Retroviruses 17:161-168(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=97ZA012;  
RA Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,  
RA Allen S., Shaw G.M., Hahn B.H., Gao F.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF286227; AAK30993.1; -;

DR InterPro: IPR000012; HIV\_OREXR.  
DR Pfam: PF00522; VPR; 1.  
DR PRINTS: PR00444; HIVPRVFX.  
SQ SEQUENCE 96 AA; 11415 MW; 839CB1B099C059B CRC64;

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 TGEVYIR 17  
|||||  
Db 55 TGEVYIR 62

RESULT 5  
ID Q99BN5 PRELIMINARY; PRT; 96 AA.  
AC Q99BN5.  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE VPR PROTEIN.  
GN VPR.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TW010-25;  
RX MEDLINE=21322026; PubMed=11429118;  
RA Scriba T.J., Treunicht F.K., Zeller M., Engelbrecht S.,  
van Rensburg E.J.,  
RT "Characterization and phylogenetic analysis of South African HIV-1  
RT subtype C accessory genes."  
RL AIDS Res. Hum. Retroviruses 17:775-781(2001).  
DR EMBL: AF325755; AAK09162.1; -  
DR InterPro: IPR000012; HIV\_OREXR.  
DR Pfam: PF00522; VPR; 1.  
DR PRINTS: PR00444; HIVPRVFX.  
SQ SEQUENCE 96 AA; 11450 MW; 663D5ED56DED0447 CRC64;

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 TGEVYIR 17  
|||||  
Db 55 TGEVYIR 62

RESULT 6  
ID Q9BPL0 PRELIMINARY; PRT; 731 AA.  
AC Q9BPL0.  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE FTZ-F1.  
GN FTZ-F1.  
OS Schistosoma mansoni (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
OC Schistosomatidae; Schistosomatidae; Schistosoma.  
OX NCBI\_TaxID=6183;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mendonca R.L., Bouton D., Vanacker J.M., Laudet V., Pierce R.;  
RT "Cloning and functional characterization of a Schistosoma mansoni  
RT homologue of the FTZ-F1 nuclear receptor."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.

DR EMBL: AF158103; AAG49449.1; -  
DR HSSP: P03372; IHCO.  
DR InterPro: IPR000536; Hormone\_rec\_11g.  
DR InterPro: IPR001723; Strdhormone\_receptor.  
DR InterPro: IPR001628; zf-C4.  
DR Pfam: PF00104; hormone\_rec\_1.  
DR Pfam: PF00105; zf-C4; 1.  
DR PRINTS: PR00398; STRDHORMONER.  
DR PRINTS: PR00047; STROIDFINGER.  
DR SMART: SM00430; HOLI; 1.  
DR SMART: SM00399; ZNF\_C4; 1.  
DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; UNKNOWN.1.  
KW DNA-binding; Nuclear protein; Receptor; transcription regulation;  
KW Zinc-finger.  
SQ SEQUENCE 731 AA; 78130 MW; 20129AF9AAF30175 CRC64;

Query Match  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 TSNNSNL 75  
|||||  
Db 714 TSNNSNL 721

RESULT 7  
ID O9K395 PRELIMINARY; PRT; 1251 AA.  
AC O9K395.  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE TYPE E BOTULINUM TOXIN.  
GN BONT/E.  
OS Clostridium butyricum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1492;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL 095;  
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
Karasawa T.;  
RT "C. butyricum (LCL 095) gene for type E botulinum toxin."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL 155 (KZ 1885);  
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,  
Kato H., Nakamura S., Karasawa T.;  
RT "C. butyricum (LCL 155) gene for type E botulinum toxin."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KZ 1899;  
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
Karasawa T.;  
RT "C. butyricum (KZ 1899) gene for type E botulinum toxin."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KZ 1897;  
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
Karasawa T.;  
RT "C. butyricum (KZ 1897) gene for type E botulinum toxin."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KZ 1898;  
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
Karasawa T.;  
RT "C. butyricum (KZ 1898) gene for type E botulinum toxin."

```

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-K2 1886;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RT "C. butyricum (K2 1886) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN-K2 1887;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RT "C. butyricum (K2 1887) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN-K2 1889;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RT "C. butyricum (K2 1889) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN-K2 1890;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RT "C. butyricum (K2 1890) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN-K2 1891;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RT "C. butyricum (K2 1891) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-LCL 063;
RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RT "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB037714; BAB03522.1; -
DR EMBL: AB037704; BAB03512.1; -
DR EMBL: AB037705; BAB03513.1; -
DR EMBL: AB037706; BAB03514.1; -
DR EMBL: AB037707; BAB03515.1; -
DR EMBL: AB037708; BAB03516.1; -
DR EMBL: AB037709; BAB03517.1; -
DR EMBL: AB037710; BAB03518.1; -
DR EMBL: AB037711; BAB03519.1; -
DR EMBL: AB037712; BAB03520.1; -
DR EMBL: AB037713; BAB03521.1; -
DR HSSP: P10845; 3BTA.
DR MEROPS: M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOMN_1.
SQ SEQUENCE 1251 AA; 143751 MW; 2021FAE427070296 CRC64;

```

```

Query Match 5.6%; Score 8; DB 2; Length 1251;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 85 GNCTMNF 92
DB 1193 GNCTMNF 1200

```

```

RESULT 8
Q9FAR6 PRELIMINARY; PRT; 1255 AA.
ID Q9FAR6;
AC Q9FAR6;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE TYPE E BOTULINUM TOXIN.
GN BONT/E.
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
ON NCBI_TaxID=1492;
RX MEDLINE=20509829; PubMed=11055954;
RP SEQUENCE FROM N.A.
RC STRAIN-BL 6340/ATCC 43755/BL 5520/K2 147;
RA Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RT "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
butyricum Strains.";
RL Appl. Environ. Microbiol. 66:4992-4997(2000).
DR HSSP: P10845; 3BTA.
DR InterPro: IPR000395; Bontoxilysin.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOMN_1.
SQ SEQUENCE 1255 AA; 143918 MW; 1B557B9D85CDE84D CRC64;

```

```

Query Match 5.6%; Score 8; DB 2; Length 1255;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 85 GNCTMNF 92
DB 1197 GNCTMNF 1204

RESULT 9
ID Q92C93 PRELIMINARY; PRT; 67 AA.
AC Q92C93;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE L1N1298 PROTEIN.
GN L1N1298.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
ON NCBI_TaxID=1642;
RX PubMed=11679669;
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RT Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Darvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Etienne K.-D., Fajl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluteter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).

```



DR EMBL: AL596168; CAC96529.1; -  
 DR Listlist: L1N01298; -  
 KW Complete Proteome.  
 SQ SEQUENCE 67 AA; 7875 MW; E2E37AE3BD0F3E4 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 39 INVDRO 45  
 DB 31 INVVDRO 37

RESULT 10  
 ID 042077 PRELIMINARY; PRT; 96 AA.  
 AC 042077;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VPR PROTEIN.  
 GN VPR.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Song J., Wang B., Ge Y.C., Dwyer D., Dowton D., Cunningham A.,  
 RL Submitted (SFP-1997) to the EMBL/GenBank/DBD databases.  
 DR EMBL: AF000338; AAB70135.1; -  
 DR EMBL: AF000334; AAB70133.1; -  
 DR EMBL: AF000336; AAB70133.1; -  
 DR InterPro: IPR000012; HIV\_ORFXR.  
 DR Pfam: PF00522; VPR.1.  
 DR PRINTS: PR00444; HIVPRVFX.  
 SQ SEQUENCE 96 AA; 11365 MW; 99DCA04651392AF CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 GVEYIR 17  
 DB 56 GVEYIR 62

RESULT 11  
 ID 089653 PRELIMINARY; PRT; 96 AA.  
 AC 089653;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE VPR PROTEIN (FRAGMENT).  
 GN VPR.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-MOTHER PAIR D;  
 RX MEDLINE=98325222; PubMed=9658150;  
 RA Yedavalli V.R., Chappey C., Ahmad N.;  
 RT "Maintenance of an intact human immunodeficiency virus type 1 vpr gene  
 following mother-to-infant transmission.";  
 RL J. Virol. 72:6937-6943(1998).  
 DR EMBL: AF042966; AAC41130.1; -  
 DR InterPro: IPR000012; HIV\_ORFXR.  
 DR Pfam: PF00522; VPR.1.

DR PRINTS: PR00444; HIVPRVFX.  
 FT NON\_TER 96  
 SQ SEQUENCE 96 AA; 11320 MW; 5995598CDF0E7FBD CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 GVEYIR 17  
 DB 56 GVEYIR 62

RESULT 12  
 ID 089654 PRELIMINARY; PRT; 96 AA.  
 AC 089654;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE VPR PROTEIN (FRAGMENT).  
 GN VPR.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-MOTHER PAIR D;  
 RX MEDLINE=98325222; PubMed=9658150;  
 RA Yedavalli V.R., Chappey C., Ahmad N.;  
 RT "Maintenance of an intact human immunodeficiency virus type 1 vpr gene  
 following mother-to-infant transmission.";  
 RL J. Virol. 72:6937-6943(1998).  
 DR EMBL: AF042966; AAC41131.1; -  
 DR InterPro: IPR000012; HIV\_ORFXR.  
 DR Pfam: PF00522; VPR.1.  
 DR PRINTS: PR00444; HIVPRVFX.  
 FT NON\_TER 96  
 SQ SEQUENCE 96 AA; 11320 MW; 5995598CDF0E7FBD CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 GVEYIR 17  
 DB 56 GVEYIR 62

RESULT 13  
 ID 089670 PRELIMINARY; PRT; 96 AA.  
 AC 089670;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE VPR PROTEIN (FRAGMENT).  
 GN VPR.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-INFANT PAIR D;  
 RX MEDLINE=98325222; PubMed=9658150;  
 RA Yedavalli V.R., Chappey C., Ahmad N.;  
 RT "Maintenance of an intact human immunodeficiency virus type 1 vpr gene  
 following mother-to-infant transmission.";  
 RL J. Virol. 72:6937-6943(1998).  
 DR EMBL: AF042968; AAC41147.1; -  
 DR InterPro: IPR000012; HIV\_ORFXR.

DR Pfam: PF00522; VPR: 1.  
 DR PRINTS: PR00444; HIVPRVPX.  
 FT NON\_TER 96  
 SQ SEQUENCE 96 AA; 11329 MM; 8DA5598CDF0E672C CRC64;

Query Match 4.9%; Score 7; DB 15; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GVEYIIR 17  
 |||||  
 Db 56 GVEYIIR 62

RESULT 14  
 O89677  
 ID O89677 PRELIMINARY; PRT; 96 AA.

AC O89677;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE VPR PROTEIN (FRAGMENT).  
 GN VPR.

OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_Taxid=11676;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-INFANT PAIR D;  
 RX MEDLINE=98325222; PubMed=9658150;  
 RA Yedavalli V.R., Chappey C., Ahmad N.;  
 RT Maintenance of an intact human immunodeficiency virus type 1 vpr gene  
 following mother-to-infant transmission.";  
 RL J. Virol. 72:6937-6943(1998).  
 DR EMBL: AF042990; AAC41154.1; -;  
 DR InterPro: IPR000012; HIV\_ORFPR.

DR Pfam: PF00522; VPR: 1.  
 DR PRINTS: PR00444; HIVPRVPX.  
 FT NON\_TER 96  
 SQ SEQUENCE 96 AA; 11320 MM; 5995598CDF0E7FBD CRC64;

Query Match 4.9%; Score 7; DB 15; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GVEYIIR 17  
 |||||  
 Db 56 GVEYIIR 62

RESULT 15  
 O89678

ID O89678 PRELIMINARY; PRT; 96 AA.

AC O89678;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE VPR PROTEIN (FRAGMENT).  
 GN VPR.

OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_Taxid=11676;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-INFANT PAIR D;  
 RX MEDLINE=98325222; PubMed=9658150;  
 RA Yedavalli V.R., Chappey C., Ahmad N.;  
 RT Maintenance of an intact human immunodeficiency virus type 1 vpr gene  
 following mother-to-infant transmission.";  
 RL J. Virol. 72:6937-6943(1998).  
 DR EMBL: AF042991; AAC41155.1; -;

DR InterPro: IPR000012; HIV\_ORFPR.  
 DR Pfam: PF00522; VPR: 1.  
 DR PRINTS: PR00444; HIVPRVPX.  
 FT NON\_TER 96  
 SQ SEQUENCE 96 AA; 11320 MM; 5995598CDF0E7FBD CRC64;

Query Match 4.9%; Score 7; DB 15; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 GVEYIIR 17  
 |||||  
 Db 56 GVEYIIR 62


Search completed: August 15, 2002, 11:24:10  
 Job time: 697 sec

Thu Aug 15 12:38:24 2002

us-08-981-087a-4.rpt

---

Page 7





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:33 ; Search time 80.39 seconds  
(without alignments)  
927.489 Million cell updates/sec

Title: US-08-981-087a-1

Sequence: 1 STTNDKILILYENKLYKKIK.....TSSNCCFWSFTSKHEHMOEN 431

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size: 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_YEAST:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	431	100.0	1278	2	057236
2	31	7.2	1280	2	092A5
3	22	5.1	1268	2	045851
4	15	3.5	1251	2	09K395
5	15	3.5	1255	2	09PAR6
6	11	2.6	367	2	045862
7	11	2.6	367	2	045861
8	11	2.6	1296	2	045894
9	8	1.9	96	15	0994N3
10	8	1.9	96	15	0994N5
11	8	1.9	361	2	045846
12	8	1.9	361	2	045848
13	8	1.9	441	2	09X708
14	8	1.9	540	16	09A5Z8
15	8	1.9	731	5	09BPLO
16	8	1.9	1072	16	09CF64

17	8	1.9	1291	2	092A58	092A58 Clostridium
18	8	1.9	1291	2	008077	008077 Clostridium
19	8	1.9	1291	2	093671	093671 Clostridium
20	8	1.9	1291	2	0933K0	0933K0 Clostridium
21	8	1.9	1291	2	090W19	090W19 Clostridium
22	7	1.6	67	16	092C93	092C93 Clostridium
23	7	1.6	67	16	092C93	092C93 Clostridium
24	7	1.6	67	16	092C93	092C93 Clostridium
25	7	1.6	67	16	092C93	092C93 Clostridium
26	7	1.6	67	16	092C93	092C93 Clostridium
27	7	1.6	67	16	092C93	092C93 Clostridium
28	7	1.6	67	16	092C93	092C93 Clostridium
29	7	1.6	67	16	092C93	092C93 Clostridium
30	7	1.6	67	16	092C93	092C93 Clostridium
31	7	1.6	67	16	092C93	092C93 Clostridium
32	7	1.6	67	16	092C93	092C93 Clostridium
33	7	1.6	67	16	092C93	092C93 Clostridium
34	7	1.6	67	16	092C93	092C93 Clostridium
35	7	1.6	67	16	092C93	092C93 Clostridium
36	7	1.6	67	16	092C93	092C93 Clostridium
37	7	1.6	67	16	092C93	092C93 Clostridium
38	7	1.6	67	16	092C93	092C93 Clostridium
39	7	1.6	67	16	092C93	092C93 Clostridium
40	7	1.6	67	16	092C93	092C93 Clostridium
41	7	1.6	67	16	092C93	092C93 Clostridium
42	7	1.6	67	16	092C93	092C93 Clostridium
43	7	1.6	67	16	092C93	092C93 Clostridium
44	7	1.6	67	16	092C93	092C93 Clostridium
45	7	1.6	67	16	092C93	092C93 Clostridium

## ALIGNMENTS

RESULT 1  
ID 057236 PRELIMINARY: PRT: 1278 AA.  
AC 057236; 045863;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE BOTULINUM NEUROTOXIN TYPE F (BONT/E PROTEIN).  
GN BONT/F.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium  
OX NCBI\_TaxID=1491;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-NCIC 10281;  
RA Hutson R.A., Collins M.D.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN-NCIC 1028;  
RX MEDLINE=94013372; PubMed=8408542;  
RA Campbell K., East A.K., Collins M.D.;  
RT "Gene probes for identification of the botulinum neurotoxin gene and specific identification of neurotoxin types B, E, and F.";  
RL J Clin. Microbiol. 31:2255-2262(1993).  
RN (4)  
RP SEQUENCE OF 1-27 FROM N.A.  
RC STRAIN-NCIC 1028;  
RX MEDLINE=96404102; PubMed=9732534;  
RA East A.K., Bhandari M., Hleim S., Collins M.D.;  
RT "Analysis of the botulinum neurotoxin type F gene clusters in proteolytic and nonproteolytic Clostridium botulinum and Clostridium bartlettii.";  
RL Curr. Microbiol. 37:262-268(1998).

X77082  
L35494c

DR EMBL: X6444; CA57358.1; - 8/15/02 - Seq 1-4  
 DR EMBL: X6444; CA57358.1; - 8/15/02 - Seq 1-4  
 DR EMBL: X70821; CA50152.1; - 10/18  
 DR EMBL: X69564; CA67512.1; - 10/18  
 DR HSSP: P10845; 3BTA; - 10/18  
 DR MEROPS: M27\_002; -  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_Mpeptidase.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOTOXILYSIN.  
 DR PRODOM: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 DR Neurotoxin.  
 KW SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64; 848-1001

Query Match 100.0%; Score 431; DB 2; Length 1278;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 SYNDKILILFYFNKLYKKIKDNDILDMRYENKFDISGYSNISINGDYIYSTNRQF 60  
 DB 848 SYNDKILILFYFNKLYKKIKDNDILDMRYENKFDISGYSNISINGDYIYSTNRQF 907  
 QY 61 GYSSKSEVNIQNDIIVNGRYONFSISFWRIKPKFNKYNLNNETIIDCIANNNSG 120  
 DB 908 GYSSKSEVNIQNDIIVNGRYONFSISFWRIKPKFNKYNLNNETIIDCIANNNSG 967  
 QY 121 WKISLNTKIIWTLQDFAGNNOQLVFNTOYOMISISDYINKMIFVITNNRLGNSRYING 180  
 DB 968 WKISLNTKIIWTLQDFAGNNOQLVFNTOYOMISISDYINKMIFVITNNRLGNSRYING 1027  
 QY 181 NLIDEKISINLGIHVSNDILFKIVGCDTRYVGIRYKRVDFELGTEIFLYSDEBDP 240  
 DB 1028 NLIDEKISINLGIHVSNDILFKIVGCDTRYVGIRYKRVDFELGTEIFLYSDEBDP 1087  
 QY 241 SILDFMGNYLLYKRYLLMLLFRDKSITONSFLNINQOQYQKRNITSNRLYGV 300  
 DB 1088 SILDFMGNYLLYKRYLLMLLFRDKSITONSFLNINQOQYQKRNITSNRLYGV 1147  
 QY 301 EVIIRKNGSTDISDNFVRKNDLAYINVDVDEYRLYADISIAKPKRIKLIKRTSSN 360  
 DB 1148 EVIIRKNGSTDISDNFVRKNDLAYINVDVDEYRLYADISIAKPKRIKLIKRTSSN 1207  
 QY 361 NSLQIIVMDSIGNNCTWNFQNNNGNIGLLGFHSNNLVASSWYNNIRKNTSSNGCFS 420  
 DB 1208 NSLQIIVMDSIGNNCTWNFQNNNGNIGLLGFHSNNLVASSWYNNIRKNTSSNGCFS 1267  
 Y 421 FISKEHGOEN 431  
 DB 1268 FISKEHGOEN 1278

RESULT 2  
 ID 09ZAU5 PRELIMINARY; PRT: 1280 AA.  
 AC 09ZAU5;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE BONT. PROTEIN.  
 GN BONT.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 3281 (ATCC 43757);  
 RC MEDLINE=98440323; PubMed=9767710;  
 RA Santos-Buelga J., Collins M.D., East A.K.;  
 RT "Characterization of the genes encoding the Botulinum neurotoxin  
 complex in a strain of Clostridium botulinum producing type B & F

RT neurotoxins.";  
 RL Curr. Microbiol. 37:312-318(1998).  
 DR EMBL: Y13631; CA73972.1; -  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27\_002; -  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_Mpeptidase.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOTOXILYSIN.  
 DR PRODOM: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 DR SEQUENCE 1280 AA; 147487 MW; D0F748976EBC222C CRC64;

Query Match 7.2%; Score 31; DB 2; Length 1280;  
 Best Local Similarity 100.0%; Pred. No. 1; 7e-21;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 SISDYINKWIFVITNNRLGNSRYINGNLI 183  
 DB 1003 SISDYINKWIFVITNNRLGNSRYINGNLI 1033

RESULT 3  
 ID 045851 PRELIMINARY; PRT: 1268 AA.  
 AC 045851;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NEUROTOXIN TYPE F.  
 GN BONT /F.  
 OS Clostridium baratii.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1561;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=93252228; PubMed=8486245;  
 RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,  
 RA Richardson P.T.;  
 RT "Nucleotide sequence of the gene coding for Clostridium baratii type F  
 neurotoxin: Comparison with other clostridial neurotoxins.";  
 RL FEMS Microbiol. Lett. 108:175-182(1993).  
 DR EMBL: X68262; CA48329.1; -  
 DR HSSP: P10845; 3BTA.  
 DR MEROPS: M27\_002; -  
 DR InterPro: IPR000395; Bontoxilysin.  
 DR InterPro: IPR000130; Zn\_Mpeptidase.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS: PR00760; BONTOTOXILYSIN.  
 DR PRODOM: PD001963; Bontoxilysin; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 DR SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 5.1%; Score 22; DB 2; Length 1268;  
 Best Local Similarity 100.0%; Pred. No. 1; 1e-12;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 DNILFKIVGCDTRYVGIRYK 219  
 DB 1036 DNILFKIVGCDTRYVGIRYK 1057

RESULT 4  
 ID 09K395 PRELIMINARY; PRT: 1251 AA.  
 AC 09K395;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TYPE E BOTULINUM TOXIN.

CN BONT/E  
 OS Clostridium butyricum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 CX NCBI\_TaxID=1492;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1886;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1886) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LCL 155 (KZ 1885);  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,  
 RA Kato H., Nakamura S., Karsawa T.;  
 RT "C. butyricum (LCL 155) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1897;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1897) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1897;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1897) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1898;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1898) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1886;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1886) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1887;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1887) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1889;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1889) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1890;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1890) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1891;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;

RT "C. butyricum (KZ 1891) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LCL 063;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (LCL 063) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1892;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1892) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1893;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1893) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1894;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1894) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1895;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1895) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1896;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1896) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [17]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1897;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1897) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [18]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1898;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1898) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [19]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1899;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;  
 RT "C. butyricum (KZ 1899) gene for type E botulinum toxin.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [20]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KZ 1900;  
 RA Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,  
 RA Karsawa T.;

Query Match 3.5%; Score 15; DB 2; Length 1251;  
 Best local similarity 100.0%; Pred. No. 8.4e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 ISDYINKWIEFTTIN 168  
 DB 983 ISDYINKWIEFTTIN 997

RESULT 5  
 Q9FAR6 PRELIMINARY; PRT; 1255 AA.  
 AC Q9FAR6;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TYPE E BOTULINUM TOXIN.  
 GN BONT/E.  
 OS Clostridium butyricum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 CX NCBI\_TaxID=1492;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;  
 RX MBDLINE=20509823; PubMed=11055954;  
 RA Wang X., Maegawa T., Karsawa T., Kozaki S., Tsukamoto K., Gyobu Y.,  
 RA Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;  
 RT "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium  
 butyricum Strains.";  
 RL Appl. Environ. Microbiol. 66:4992-4997(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AB039264; BAB12249.1; -;  
 DR HSSP; P10845; 38TA.  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTTOXILYSIN.  
 DR PRODOM; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 SO SEQUENCE 1255 AA; 143918 MW; 1B557B9D5CDB84D CRC64;

Query Match 3.5%; Score 15; DB 2; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 8,4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 ISDYINKWIFVTITN 168  
|||||  
DB 986 ISDYINKWIFVTITN 1000

RESULT 6  
O45862 PRELIMINARY; PRT; 367 AA.  
ID O45862  
AC O45862;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).  
GN BONT/E.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
NCBI\_TaxID=1491;  
OY [1]  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TYPE E, HAZEN 36208 (ATCC 9564);  
RX MEDLINE=94013372; PubMed=8408542;  
RA Campbell K., East A.K., Collins M.D.;  
RT "Gene probes for identification of the botulin neurotoxin gene and  
RT specific identification of neurotoxin types B, E, and F.";  
RL J. Clin. Microbiol. 31:2255-2262(1993).  
DR EMBL; X70815; CAA50146.1; -.  
DR HSSP; P10845; 3BTA.  
KW Neurotoxin.  
FT NON\_TER 1 1  
SQ SEQUENCE 367 AA; 42854 MW; 0810595B3A865570 CRC64;

Query Match 2.6%; Score 11; DB 2; Length 367;  
Best Local Similarity 100.0%; Pred. No. 0.026;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 NFSISFWVRIP 96  
|||||  
DB 299 NFSISFWVRIP 309

RESULT 7  
O45861 PRELIMINARY; PRT; 367 AA.  
ID O45861  
AC O45861;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).  
GN BONT/E.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
NCBI\_TaxID=1491;  
OY [1]  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TYPE E, VH (DOLMAN);  
RX MEDLINE=94013372; PubMed=8408542;  
RA Campbell K., East A.K., Collins M.D.;  
RT "Gene probes for identification of the botulin neurotoxin gene and  
RT specific identification of neurotoxin types B, E, and F.";  
RL J. Clin. Microbiol. 31:2255-2262(1993).  
DR EMBL; X70818; CAA50149.1; -.  
DR HSSP; P10845; 3BTA.  
KW Neurotoxin.  
FT NON\_TER 1 1  
SQ SEQUENCE 367 AA; 42854 MW; 0810595B3A865570 CRC64;

SQ SEQUENCE 367 AA; 42902 MW; 346A610C2FF70262 CRC64;

Query Match 2.6%; Score 11; DB 2; Length 367;  
Best Local Similarity 100.0%; Pred. No. 0.026;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 NFSISFWVRIP 96  
|||||  
DB 299 NFSISFWVRIP 309

RESULT 8  
O45894 PRELIMINARY; PRT; 1296 AA.  
ID O45894  
AC O45894; P77780;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE BOTULINUM NEUROTOXIN TYPE A (TYPE A NEUROTOXIN).  
GN BONT OR ATX.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
NCBI\_TaxID=1491;  
OY [1]  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KYOTO-F;  
RX MEDLINE=94143603; PubMed=8310180;  
RA Williams A., East A.K., Lawson P.A., Collins M.D.;  
RT "Sequence of the gene coding for the neurotoxin of Clostridium  
RT botulinum type A associated with infant botulism: comparison with  
RT other clostridial neurotoxins.";  
RL Res. Microbiol. 144:547-556(1993).  
RN [2]  
RP SEQUENCE OF 1-65 FROM N.A.  
RC STRAIN-62A;  
RX MEDLINE=97016817; PubMed=8863443;  
RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
RT "Organization and phylogenetic interrelationships of genes encoding  
RT components of the botulinum toxin complex in proteolytic Clostridium  
RT botulinum types A, B, and F: evidence of chimeric sequences in the  
RT gene encoding the nontoxic nonhemagglutinin component.";  
RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
RN [3]  
RP SEQUENCE OF 1-18 FROM N.A.  
RC STRAIN-TYPE A NIH;  
RX MEDLINE=96096783; PubMed=8521962;  
RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
RT components of Clostridium botulinum type A progenitor toxins.";  
RL FEBS Lett. 376:41-44(1995).  
DR EMBL; X73423; CAA51824.1; -.  
DR EMBL; X92973; CAA63551.1; -.  
DR EMBL; X89274; CAA61234.1; -.  
DR EMBL; D67030; BAA11051.1; -.  
DR HSSP; P10845; 3BTA.  
DR InterPro; IPR000395; Bontoxilysin.  
DR Pfam; PF01742; Peptidase\_M27; 1.  
DR PRINTS; PR00760; BONTTOXILYSIN.  
DR Prodom; PD001963; Bontoxilysin; 1.  
KW Neurotoxin.  
SQ SEQUENCE 1296 AA; 149410 MW; 6F12E7BF28ED69D1 CRC64;

Query Match 2.6%; Score 11; DB 2; Length 1296;  
Best Local Similarity 100.0%; Pred. No. 0.073;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 161 WIFVTITNRL 171  
|||||  
DB 1014 WIFVTITNRL 1024



```

RESULT 9
ID 0994N3 PRELIMINARY: PRT: 96 AA.
AC 0994N3;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE VPR PROTEIN.
GN VPR.
OS Human immunodeficiency virus type 1.
OC Viruses: Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=97A012;
RA Roderburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA Allen S., Shaw G.M., Hahn B.H., Gao F.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF286227; AK30993.1;
DR InterPro: IPR000012; HIV_ORFXR.
DR Pfam: PF00522; VPR.1.
DR PRINTS: PR00444; HIVVPRPX.
SQ SEQUENCE 96 AA; 11415 MW; 839CB1B0999C059B CRC64;

Query Match 1.9%; Score 8; DB 15; Length 96;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 TGVEVIIR 305
DB 55 TGVEVIIR 62

RESULT 10
ID 099BN5 PRELIMINARY: PRT: 96 AA.
AC 099BN5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE VPR PROTEIN.
GN VPR.
OS Human immunodeficiency virus type 1.
OC Viruses: Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TV010-25;
RA MEDLINE=21322026; PubMed=11429118;
RA Scriba T.J., Treurnicht F.K., Zeller M., Engelbrecht S.,
RA van Rensburg E.J.;
RT "Characterization and phylogenetic analysis of South African HIV-1
RT subtype C accessory genes.";
RL AIDS Res. Hum. Retroviruses 17:775-781(2001).
DR EMBL: AF325755; AK09162.1;
DR InterPro: IPR000012; HIV_ORFXR.
DR Pfam: PF00522; VPR.1.
DR PRINTS: PR00444; HIVVPRPX.
SQ SEQUENCE 96 AA; 11450 MW; 663D5ED56DD0447 CRC64;

Query Match 1.9%; Score 8; DB 15; Length 96;

```

```

Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 TGVEVIIR 305
DB 55 TGVEVIIR 62

RESULT 11
ID 045846 PRELIMINARY: PRT: 361 AA.
AC 045846;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE B, NON-PROTEOLYTIC 2129B (SCOTT);
RA MEDLINE=9401372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL: X70814; CAA50145.1;
DR HSSP: P10845; 3B7A.
KW Neurotoxin.
FT NON_TER 1 1
SQ SEQUENCE 361 AA; 42175 MW; 533EA98735CD98E1 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 361;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 NNSGWKIS 124
DB 325 NNSGWKIS 332

RESULT 12
ID 045848 PRELIMINARY: PRT: 361 AA.
AC 045848;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN BONT/B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE B, NON-PROTEOLYTIC EKUND 2B (COLWORTH 229);
RA MEDLINE=9401372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
DR EMBL: X70814; CAA50150.1;
DR HSSP: P10845; 3B7A.
KW Neurotoxin.
FT NON_TER 1 1
SQ SEQUENCE 361 AA; 42131 MW; A2E0FFC81F9533D CRC64;

```

Query Match 1.9%; Score 8; DB 2; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 NNSGKIS 124  
 |||||

DB 325 NNSGKIS 332

RESULT 13

OY 09X708 PRELIMINARY; PRT; 441 AA.  
 AC 09X708;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).  
 BONT/B.  
 Clostridium botulinum.  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 Clostridium.  
 NCBI\_TaxID=1491;  
 OX  
 RN  
 RP  
 RX MEDLINE=9343691; PubMed=10413679;  
 RA Lailli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,  
 RA Schiavo G.;  
 RT "Functional characterization of tetanus and botulinum neurotoxins  
 binding domains";  
 RL J. Cell Sci. 112:2715-2724(1999).  
 DR EMBL; AJ242628; CAB43706.1; -.  
 DR HSSP; P10845; 3BTA.  
 KW Neurotoxin.  
 FT NON TER 1 1  
 FT SEQUENCE 441 AA; 52772 MW; 721D0B468E8C95N4 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 441;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 NNSGKIS 124  
 |||||

DB 116 NNSGKIS 123

RESULT 14

OY 09A928 PRELIMINARY; PRT; 540 AA.  
 AC 09A928;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN CC0813.  
 GN CC0813.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RN  
 RP  
 RX STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.D., Durkin A.S., Gwin M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty K., Berry K.,  
 RA Uitterback T., Tran K., Wolf A., Vanathavan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro J.C., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005758; AAK22798.1; -.  
 DR TIGR; CC0813; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 540 AA; 59648 MW; 72BC45442BEF99FD CRC64;

Query Match 1.9%; Score 8; DB 16; Length 540;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 238 PDPSTLKD 245  
 |||||

DB 71 PDPSTLKD 78

RESULT 15

OY 09BPL0 PRELIMINARY; PRT; 731 AA.  
 AC 09BPL0;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE FTZ-F1.  
 GN FTZ-F1.  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeididae;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxID=6183;  
 RN  
 RP  
 RX SEQUENCE FROM N.A.  
 RA Mendonca R.L., Bouton D., Vanacker J.-M., Laudet V., Pierce R.;  
 RT "Cloning and functional characterization of a Schistosoma mansoni  
 homologue of the FTZ-F1 nuclear receptor";  
 RL submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL; AF158103; AAG49449.1; -.  
 DR HSSP; P03372; 1HCO.  
 DR InterPro: IPR000536; Hormone\_rec\_1lg.  
 DR InterPro: IPR001723; Steroidhormone\_receptor.  
 DR InterPro: IPR001628; zf-C4.  
 DR Pfam; PF00104; hormone\_rec\_1.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00398; STRDHORMONER.  
 DR PRINTS; PR00047; STROIDFINGER.  
 DR SMART; SM00430; STROIDFINGER.  
 DR SMART; SM00399; znf-C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; UNKNOWN\_1.  
 DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;  
 KW Zinc-finger.  
 SQ SEQUENCE 731 AA; 78130 MW; 20129AF9AAE30175 CRC64;

Query Match 1.9%; Score 8; DB 5; Length 731;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 356 TSNSNSL 363  
 |||||

DB 714 TSNSNSL 721

Search completed: August 15, 2002, 11:24:06  
 Job time: 693 sec

---



5